



Universidade de Aveiro
Ano 2012/2013

Departamento de Química

João Nuno Lágio
Bordadágua

**Descoberta de novas suberinases através de
uma abordagem metagenómica**



Universidade de Aveiro
Ano 2012/2013

Departamento de Química

João Nuno Lágio
Bordadágua

**Descoberta de novas suberinases através de
uma abordagem metagenómica**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biotecnologia, ramo em Biotecnologia Molecular, realizada sob a orientação científica da Doutora Conceição Egas, Investigadora da Unidade de Serviços Avançados do Biocant e do Prof. Doutor Jorge Saraiva, Investigador Auxiliar do Departamento de Química da Universidade de Aveiro.

O júri

Presidente

Prof. Doutor António Carlos Matias Correia

Professor Catedrático do Departamento de Biologia da Universidade de Aveiro

Doutora Maria Conceição Venâncio Egas

Diretora da Unidade de Serviços Avançados do Biocant

Prof. Doutor Jorge Manuel Alexandre Saraiva

Investigador Auxiliar do Departamento de Química da Universidade de Aveiro

Doutor Artur Jorge da Costa Peixoto Alves

Investigador Auxiliar do Departamento de Biologia & CESAM da Universidade de Aveiro

Agradecimentos

Inevitavelmente, dedico todo o trabalho à minha família. Por mais palavras que tente, não há como descrever a moral e apoio prestado pelos meus pais e irmão, apenas eles sabem e sentem a verdadeira união que nos liga. Sois-me TUDO.

Em segundo lugar, seria mais do que uma ofensa não agradecer à minha namorada, Clarisse Pereira. Obrigado por toda a paciência, ajuda, compreensão e carinho nesta fase da minha vida. Apenas em torno de uma mulher que se ama se pode formar uma família e se tornar um homem realizado.

Um agradecimento especial à Doutora Conceição Egas e ao Professor Doutor Jorge Saraiva pela oportunidade que me foi concebida e por todo o apoio, atenção, conhecimento e disponibilidade prestada.

Um muito obrigado a todas as pessoas que me receberam e acompanharam diariamente no Biocant, nomeadamente, Cristina Barroso, Diogo Pinho, Felipe Santos, Francisco Leisico, Hugo Froufe, Maria Simões, Miguel Monsanto, Miguel Vaz e Paula Gomes da Unidade de Serviços Avançados do Biocant.

Palavras-chave

Abordagens metagenômicas, *Crematogaster scutellaris*, degradação de cortiça, microbioma intestinal, pirosequenciação, suberinases.

Resumo

A espécie de formigas, *Crematogaster scutellaris*, nidifica na cortiça de sobreiros (*Quercus suber* L.), escavando galerias extensas que diminuem o valor comercial deste produto. Acredita-se que a degradação da cortiça é efetuada por enzimas digestivas secretadas pelas glândulas da própria formiga, como também por microrganismos simbiossantes presentes no seu sistema digestivo intestinal – o microbioma intestinal.

Tendo em conta que a cortiça é composta maioritariamente por suberina, o objetivo principal deste trabalho foi a pesquisa de genes do microbioma intestinal que codificam suberinases (cutinases, feruloil esterases e lacases). Para tal, estabeleceu-se uma estratégia de amplificação dos genes das enzimas alvo e respetiva sequenciação e anotação funcional.

Na primeira fase efetuou-se o estudo da composição taxonómica do microbioma intestinal. O DNA total da comunidade microbiana presente no intestino e estômago das formigas (metagenoma) foi extraído e as sequências correspondentes aos genes do rRNA 16S amplificadas por PCR e pirosequenciadas. Da comunidade bacteriana destaca-se a predominância do filo *Proteobacteria* com 68,0% do número total de sequências, seguindo-se os filos *Bacteroidetes* (17,2%), *Firmicutes* (7,0%), *Acidobacteria* (4,6%) e *Actinobacteria* (3,2%). Os taxa identificados foram pesquisados na literatura para a presença de genes de cutinases, feruloil esterases e lacases.

Numa segunda fase, recorrendo a bases de dados e ferramentas bioinformáticas, procedeu-se ao desenho de primers “consenso” para zonas conservadas de cutinases, feruloil esterases e lacases e as enzimas alvo foram amplificadas a partir do DNA do microbioma intestinal. As enzimas foram posteriormente pirosequenciadas para identificação e anotação funcional.

A estratégia funcionou bem para as lacases, no entanto não foi possível identificar sequências para as restantes enzimas alvo (cutinases e feruloil esterases). A anotação funcional das sequências obtidas identificou 4 lacases bacterianas pertencentes à família multi-cobre oxidase, sendo 3 anotadas como proteína A de resistência ao cobre e uma como precursor da oxidase CueO. Os genes apresentaram elevada homologia com lacases dos géneros *Rhodococcus*, *Serratia* e *Tetrasphaera*, sugerindo que se trate de novas enzimas com capacidade de oxidação de compostos fenólicos presentes na lenhina da cortiça.

Apesar dos resultados animadores, são necessários estudos subsequentes para desvendar a totalidade dos genes codificantes de lacases, e comprovar a sua especificidade e actividade enzimáticas em vários substratos fenólicos.

Keywords

Cork degradation, *Crematogaster scutellaris*, gut microbiome, metagenomic approaches, pyrosequencing, suberinases.

Abstract

The ant species, *Crematogaster scutellaris*, nests in the bark of cork oak (*Quercus suber* L.), digging extensive galleries that decrease the commercial value of this product. It is believed that cork degradation is carried out by digestive enzymes secreted from the glands of the host (ant), as well as by symbiotic microorganisms present in their intestinal digestive system – the gut microbiome.

Given that cork is composed mostly by suberin, the main objective of this work was the targeted-gene search in the gut microbiome for encoded suberinases (cutinases, feruloyl esterases and laccases). To do so, a strategy for enzyme gene amplification and respective sequencing and functional characterization was established.

First, a study of the taxonomic composition of the gut microbiome was carried out. The total DNA of the microbial community present in the intestine and stomach of ants (metagenome) was extracted and the sequences corresponding to the 16S rRNA genes amplified by PCR and pyrosequenced. In the bacterial community stands out the predominance of the phylum Proteobacteria with 68,0% of the total number of sequences, followed by the phyla Bacteroidetes (17,2%), Firmicutes (7,0%), Acidobacteria (4,6%) and Actinobacteria (3,2%). The taxa identified were surveyed in literature for the presence of the target genes.

In a second stage, consense primers for conserved regions of cutinases, feruloyl esterases and laccases were built based on database information and bioinformatics tools and the targeted enzymes amplified from the DNA of the intestinal microbiome. The enzymes were subsequently pyrosequenced for identification and functional annotation.

The strategy established was successful for laccases, however it was not possible to identify sequences for the other targeted enzymes (cutinases and feruloyl esterases). Functional annotation of sequences identified four bacterial laccase sequences belonging to the multi-copper oxidase family, three of them annotated as "copper resistance protein A" and one as "blue copper oxidase CueO precursor." The sequences had high similarity with those of the genus *Rhodococcus*, *Serratia* and *Tetrasphaera*, and appear to be new enzymes capable of oxidizing phenolic compounds present in the lignin of cork.

In spite of the encouraging results, further studies are needed to unravel the complete genes that encode laccases and verify its enzymatic activity and specificity in a broad range of phenolic compounds.

Índice

I – INTRODUÇÃO	1
1- Cortiça	2
1.1- Importância socioeconómica em Portugal	2
1.2- Composição química global	2
1.2.1- Estrutura e composição da suberina	3
2- Suberinases	6
2.1- Cutinases	7
2.2- Feruloil esterases	8
2.3- Lacases	10
3- Fontes naturais de suberinases – <i>Crematogaster scutellaris</i>	12
3.1- <i>Crematogaster scutellaris</i> e a sua atividade nos sobreiros	13
3.2- Sistemas enzimáticos de degradação de material lenhocelulósico em insetos	14
4- Estudo de comunidades microbianas	16
4.1- Metagenómica	16
4.2- Métodos de determinação da estrutura e composição de comunidades	18
4.2.1- Pirosequenciação dos genes do rRNA 16S	19
4.3- Métodos de determinação de genes alvo codificantes de enzimas	23
4.3.1- Pirosequenciação dos genes amplificados por PCR através de primers “consenso”	24
4.4- Ferramentas bioinformáticas	25
5- Objetivos do trabalho e metodologias a utilizar	27
II – MÉTODOS	28
1- Recolha das formigas – <i>Crematogaster scutellaris</i>	29
2- Dissecção do interior do gáster das formigas	30
3- Extração do DNA da comunidade microbiana	30
4- Estudo da comunidade bacteriana – Pirosequenciação e análise dos genes do rRNA 16S	32
4.1- Amplificação por PCR da região V3V4 dos genes do rRNA 16S	32
4.2- Pirosequenciação	33
4.3- Análise de dados	34

5- Determinação de genes alvo codificantes de enzimas	35
5.1- Desenho dos primers “consenso” degenerados	35
5.2- Amplificação por PCR através de primers “consenso” degenerados	39
5.3- Pirosequenciação	41
5.4- Análise de dados	43
III – RESULTADOS E DISCUSSÃO	44
1- Composição do microbioma intestinal da formiga	45
1.1- Amplificação por PCR da região V3V4 dos genes do rRNA 16S	45
1.2- Pirosequenciação e análise da riqueza microbiana	46
1.3- Identificação taxonômica bacteriana	48
1.4- Confirmação de genes alvo em bases de dados, presentes no genoma de bactérias identificadas na flora intestinal da formiga	51
2- Pesquisa e identificação de genes de suberinasas	53
2.1- Desenho dos primers “consenso” degenerados	53
2.2- Amplificação por PCR através de primers “consenso” degenerados	55
2.3- Fragmentação mecânica dos amplicões de enzimas alvo	57
3- Caracterização das enzimas alvo	58
3.1- Pirosequenciação dos amplicões gerados pelos primers “consenso”	58
3.2- Identificação taxonômica das sequências amplificadas	59
3.3- Identificação funcional das sequências amplificadas	60
3.3.1- Identificação de lacases eucariotas	68
IV – CONCLUSÕES	70
VI – BIBLIOGRAFIA	73

Lista de Figuras

Figura 1- Representação esquemática da estrutura da suberina na cortiça do sobreiro	2
Figura 2- Exemplos de monómeros (C ₁₈) presentes no domínio alifático da suberina	3
Figura 3- Percursos fenólicos predominantes em tecidos suberosos do sobreiro	3
Figura 4- Secção transversal da parede celular de uma célula da cortiça e ampliação de uma zona da parede secundária, com destaque dos dois domínios presentes na suberina	4
Figura 5- Locais alvo de cada enzima (suberinase) na molécula de suberina	5
Figura 6- Representação da estrutura de uma cutinase com a triade-catalítica exposta (Ser, His, Asp)	7
Figura 7- Representação 3D de uma feruloil esterase e da sua triade-catalítica (Ser, His, Asp)	8
Figura 8- Representação esquemática dos átomos de cobre de uma lacase e da sua estrutura trinuclear	10
Figura 9- Imagem correspondente a uma obreira da espécie <i>Crematogaster scutellaris</i>	12
Figura 10- Danos causados pelas formigas na cortiça virgem e na cortiça de reprodução	13
Figura 11- Representação dos vários componentes do sistema digestivo da formiga	15
Figura 12- Metodologia utilizada na abordagem metagenómica que estuda todos os genes presentes numa amostra ambiental	16
Figura 13- Representação das regiões conservadas e hipervariáveis do gene rRNA 16S, correspondente a um total de 1542 pares de bases	19
Figura 14- Representação das várias etapas da pirosequenciação desde a separação da biblioteca de amplicões até à entrada das “beads” na placa de sequenciação	20

Figura 15- Série de reações enzimáticas que produzem luz durante a pirosequenciação	21
Figura 16- Metodologia utilizada na abordagem metagenômica de pesquisa de genes alvo	25
Figura 17- Localização geográfica do local onde se efetuou a recolha das formigas e imagem do “Pooter”	29
Figura 18- – Representação da composição dos primers “forward” e “reverse” e dos fragmentos amplificados dos genes do rRNA 16S	32
Figura 19- Procedimento realizado para a construção de primers “consenso” degenerados	38
Figura 20- Sumário do procedimento realizado para a biblioteca de enzimas alvo	42
Figura 21- Verificação do tamanho e qualidade do amplicão obtido para a região V3V4 dos genes do rRNA 16S	45
Figura 22- Curva de rarefação com as respetivas OTUs e sequências de alta qualidade obtidas	47
Figura 23- Análise da distribuição taxonómica da comunidade bacteriana com base nas sequências dos genes do rRNA 16S	49
Figura 24- Análise da distribuição taxonómica a nível do filo Proteobacteria, com base nas sequências dos genes do rRNA 16S	50
Figura 25- Exemplo de uma região “consenso” da enzima lacase identificada no software MEGA, através de um alinhamento múltiplo	54
Figura 26 – Representação esquemática dos primers utilizados nas 5 amplificações previstas para as enzimas alvo, assim como o tamanho esperado nos amplicões	55
Figura 27- Gel de agarose obtido após amplificação por PCR das diferentes enzimas alvo	56
Figura 28- Perfil de fragmentação das enzimas alvo obtido por ultrasonicação	57
Figura 29- Anotação taxonómica prevista pela “pipeline” de análise MG-RAST para as sequências de elevada qualidade	59
Figura 30- Anotação de sequências correspondentes a lacases bacterianas	63-64
Figura 31- Representação de árvores filogenéticas para as lacases bacterianas	65
Figura 32- Representação de PSI-Blast para as lacases bacterianas	67
Figura 33- Anotação de sequências correspondentes a lacases eucariotas	69

Lista de Tabelas

Tabela 1- Sequências dos primers utilizados na amplificação por PCR da região V3V4 do genes do rRNA 16S	32
Tabela 2- Lista dos primers “consenso” degenerados construídos para cutinases, feruloil esterases e lacases	37
Tabela 3- Números e percentagens de sequências e OTUs gerados no processamento de dados na pirosequenciação	46
Tabela 4- Dados gerados pela “pipeline” no processamento das sequências de rDNA, correspondentes à região V3V4 do genes do RNA ribossomal 16S	52
Tabela 5- Gama de valores previstos para cada uma das cinco amplificações realizadas para as enzimas alvo	55
Tabela 6- Análise estatística do upload efetuado na “pipeline” de análise MG-RAST	58

Lista de Abreviaturas

APS - Adenosina 5'-fosfossulfato
ARDRA - Análise por restrição de DNA ribossomal amplificado
Asp - Ácido aspártico
ATP - Adenosina trifosfato
CopA - Copper resistance protein A
CueO – Blue copper oxidase CueO
Da - Dalton
DDH - Técnica de hibridização do DNA
DGGE - Eletroforese em gel com gradiente de desnaturação
DNA - Ácido desoxirribonucleico
EMBL – European molecular biology laboratory
ENA – European nucleotide archive
emPCR - Reação de amplificação em cadeia em emulsão
FAEs - Feruloil esterases
Gli - Glicina
Gln - Glutamina
His - Histidina
ITS - Região intergênica do RNA ribossomal de fungos
IUB - International Union of Biochemistry
NCBI - National Center for Biotechnology Information
OTU - Unidade operacional taxonômica
Pb - Pares de bases
PCR – Amplificação por reação em cadeia da polimerase
PIB - Produto interno bruto
PPi - Pirofosfato inorgânico
RISA - Técnica de polimorfismo do espaço intergênico ribossomal
RMN - Espectroscopia de ressonância magnética nuclear
RNA - Ácido ribonucleico
rRNA - Ácido ribonucleico ribossomal
SDS-PAGE – Eletroforese em gel de poliacrilamida com dodecil-sulfato de sódio
Ser – Serina

SIB - Swiss Institute of Bioinformatics

TAT - Twin-arginine translocation pathway

TGGE - Eletroforese em gel com gradiente de temperatura

Tir – Tirosina

TM – Temperatura de melting

T-RFLP – Técnica de polimorfismo do tamanho do fragmento de restrição

W – Watt

WGS – Whole genome shotgun

I - INTRODUÇÃO

1. Cortiça

1.1 Importância socioeconómica em Portugal

A indústria da cortiça tem uma enorme relevância para o motor da economia nacional, tanto a nível da criação e manutenção de empresas, como a nível da empregabilidade e exportação. Conjuntamente, simboliza o sustento e o desenvolvimento de dois distritos, nomeadamente, Aveiro e Setúbal [1]. A estas duas regiões corresponde 87% do total de emprego gerado pela indústria corticeira, mais especificamente 75% em Aveiro e 12% em Setúbal [1].

Por ano, a indústria corticeira produz aproximadamente 300.000 toneladas de cortiça. Portugal é o líder mundial deste setor, detendo mais de 150.000 toneladas anuais (~53%) [1]. Dados referentes ao ano de 2011 mostram que o setor corticeiro representa 1% do PIB [2], o que torna a cortiça uma matéria-prima valiosa, destacada por ser usada a nível industrial para a produção de rolhas de cortiça [3]. Segundo o Ministério do Trabalho e da Solidariedade Social, em 2008, estavam em funcionamento 700 empresas, na indústria corticeira, que produziam uma média de 40 milhões de rolhas por dia, empregando cerca de dez mil operários [1]

A cortiça disponibiliza inúmeras aplicações em diferentes setores industriais, das quais originam grandes quantidades de desperdício, essencialmente pó de cortiça, ou seja, frações de baixa granulometria sem interesse comercial [4]. A modificação química das substâncias naturais ou dos compostos extraídos da cortiça possibilitará uma melhor exploração deste produto florestal, uma vez que apresenta propriedades únicas e tem grande potencial de ser transformado em compostos de valor acrescentado [4; 5].

1.2 Composição química global

A cortiça representa o tecido vegetal que reveste o tronco e os ramos do sobreiro, *Quercus suber* L., sendo constituída por células mortas ordenadas (alinhas) em parênquima suberificado, originadas pelo meristema súbero-felodérmico, também designado como felogene [6]. Este tecido possui a capacidade de divisão celular, originando células de cortiça para o exterior (súber), em maior quantidade, e células semelhantes às da madeira para o interior (feloderme), em quantidades inferiores [6].

A cortiça é um biopolímero que desempenha um papel vital na proteção do sobreiro contra alterações climáticas e agentes patogénicos, assim como, no controlo de temperatura e perda de água [7]. A suberina é o seu principal constituinte químico (~40%), seguindo-se a lenhina (~25%) e os polissacarídeos celulose e hemicelulose (~20%) [7]. A restante percentagem corresponde a extratáveis (~14%), com destaque para as ceras e os taninos; e a cinzas (~1%) [7]. Contudo, apesar do estudo da composição química da cortiça ter sido iniciado no séc. XVII, os valores médios indicados podem apresentar uma enorme discrepância, uma vez que a composição da cortiça é induzida por diversos fatores como: tipo de solo, de cortiça e de exploração florestal, clima, idade do sobreiro e diferentes metodologias de análise [6; 7; 8].

1.2.1 Estrutura e composição da suberina

A suberina fica alojada nas paredes celulares vegetais, conferindo impermeabilidade e propriedades notáveis às células de cortiça [9]. A nível macromolecular, a sua estrutura e composição ainda não estão totalmente esclarecidas [7; 9]. Com base em vários estudos anteriores, a suberina é descrita como um poliéster alifático, semelhante à cutina, ligado covalentemente a uma matriz fenólica, idêntica à lenhina, apresentando deste modo dois domínios distintos: um alifático e outro aromático (Figura 1) [5; 6; 9; 10; 11].

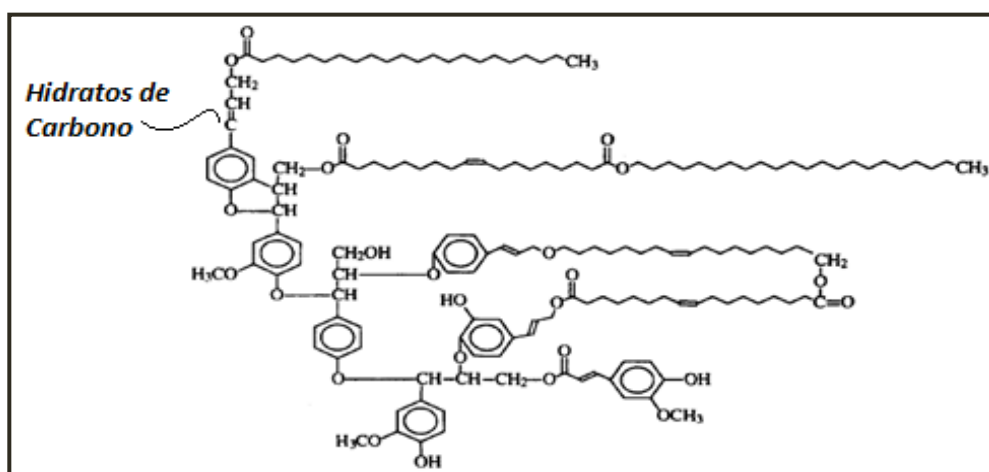


Figura 1 – Representação esquemática da estrutura da suberina na cortiça do sobreiro adaptado [11].

O domínio alifático da suberina consiste numa rede complexa de α,ω -ácidos dicarboxílicos, ω -hidroxiácidos, ácidos gordos, álcoois e glicerol [5; 7; 10]. Está localizado maioritariamente na parede secundária [12], e apresenta cadeias de carbono longas com ligações éster, cujos monómeros mais abundantes são hidroxiácidos e ácidos dicarboxílicos (C_{18}) (Figura 2) [7; 13].

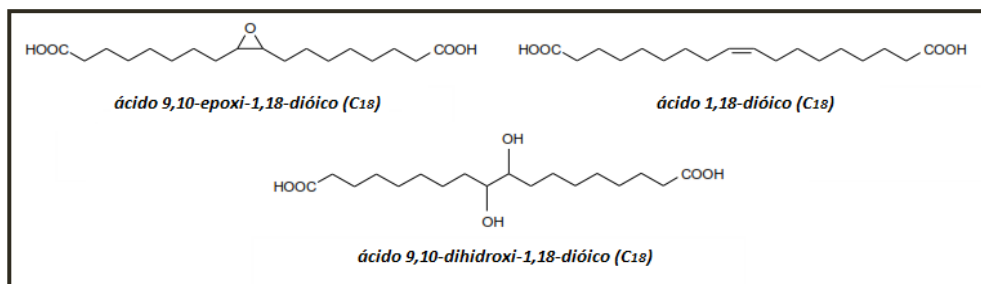


Figura 2 – Exemplos de monómeros (C_{18}) presentes no domínio alifático da suberina adaptado [13].

Por sua vez, o domínio aromático é caracterizado por uma quantidade significativa de ácidos hidroxicinâmicos, principalmente ácido ferúlico (Figura 3-A); e uma pequena quantidade de monómeros de lenhina, comparativamente a tecidos lenhificados [9; 14]. Recentemente, foram sugeridos dois tipos de compostos aromáticos para tecidos suberosos: um domínio composto maioritariamente por ácidos hidroxicinâmicos não lenhificados, localizado na parede celular secundária; e um segundo tipo de estrutura poliaromática, localizada na parede celular primária e terciária, que consiste numa lenhina do tipo guaiaquil com monómeros de álcool coniferil (Figura 3-B) [9].

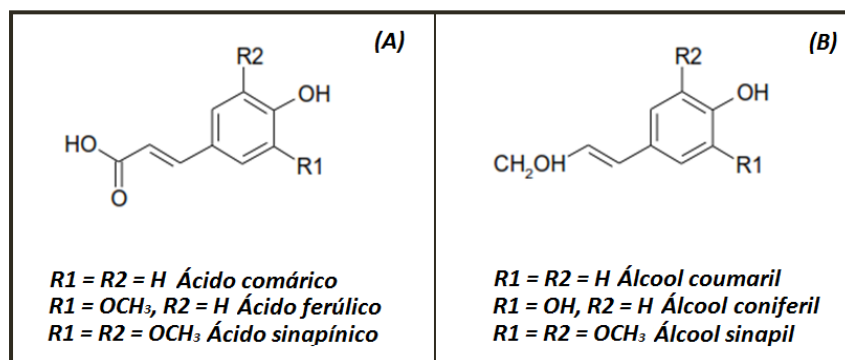


Figura 3 – Precursores fenólicos predominantes em tecidos suberosos adaptado [13].

A tentativa de demonstração da estrutura macromolecular da suberina provém de duas fontes principais: oligómeros obtidos por despolimerização parcial e observações do biopolímero intato por espectroscopia de RMN do estado sólido [5; 6; 9; 10]. Este conhecimento permitiu a elaboração de um modelo para esta estrutura (Figura 4) [9].

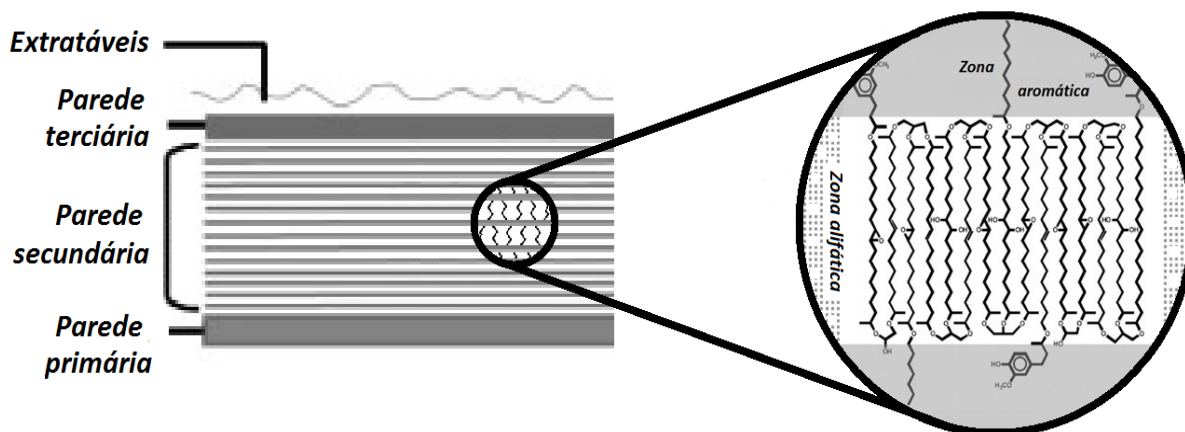


Figura 4 – Seção transversal da parede celular de uma célula da cortiça (à esquerda) e ampliação de uma zona da parede secundária, com destaque dos dois domínios presentes na suberina (à direita) adaptado [9].

A matriz lenhocelulósica, presente na parede primária e terciária, encontra-se ligada ao domínio aromático da suberina, ao qual se ligam por sua vez as cadeias longas da zona zona alifática [6; 9].

Na parede secundária vegetal, o poliéster alifático da suberina corresponde às lamelas translúcidas, enquanto as lamelas opacas representam a zona aromática. Na interface de lamelas, a esterificação da estrutura alifática com glicerol e ácido ferúlico asseguram as ligações entre a zona poliéster e os compostos aromáticos [9]. Apesar da representação estar em duas dimensões neste modelo, estas estruturas estão também na base do desenvolvimento de uma rede polimérica tridimensional [9].

Para além disso, os monómeros de cadeia longa estão esticados e orientados perpendicularmente ao plano da lamela, podendo também encontrar-se ligações intra e intermoleculares no meio destas cadeias, mais especificamente entre os grupos substituídos que contêm oxigénio [9].

2. Suberinases

Entende-se por suberinases um grupo de enzimas com capacidade de hidrolizar poliésteres naturais, como a suberina, a compostos funcionais de menor peso molecular, isto é, degradam ou despolimerizam parcialmente tecidos suberosos, convertendo-os em oligómeros ou monómeros de valor acrescentado [15].

Este conjunto de enzimas está muito pouco descrito na literatura, dado que até agora apenas está caracterizada uma esterase proveniente de *Streptomyces scabiei*, um agente patogénico da casca da batata [16]. Outro estudo relata o desenvolvimento de um fungo, *Chrysonilia sitophila*, nas pranchas de cortiça durante o processo de fabrico de rolhas, que apresenta atividade lipolítica na estrutura da suberina [17].

Em geral, o termo suberinase é intimamente associado a cutinase, dado que algumas cutinases apresentam, simultaneamente, atividade enzimática nos biopolímeros cutina e suberina [16]. Todavia, esta atividade verifica-se exclusivamente no domínio alifático da molécula da suberina [16; 17]. Desta forma, para se obter uma hidrólise eficiente da suberina é necessário que esta seja efetuada por mais do que uma enzima, especificamente, algumas suberinases terão como alvo a degradação da zona alifática, enquanto outras a da zona aromática.

Nos próximos tópicos serão caracterizadas três enzimas com potencial na degradação da suberina, mais concretamente, cutinases, feruloil esterases e lacases. Cada uma destas enzimas terá um local específico de atuação na molécula de suberina, ou seja, conforme a composição química presente em cada domínio (Figura 5).

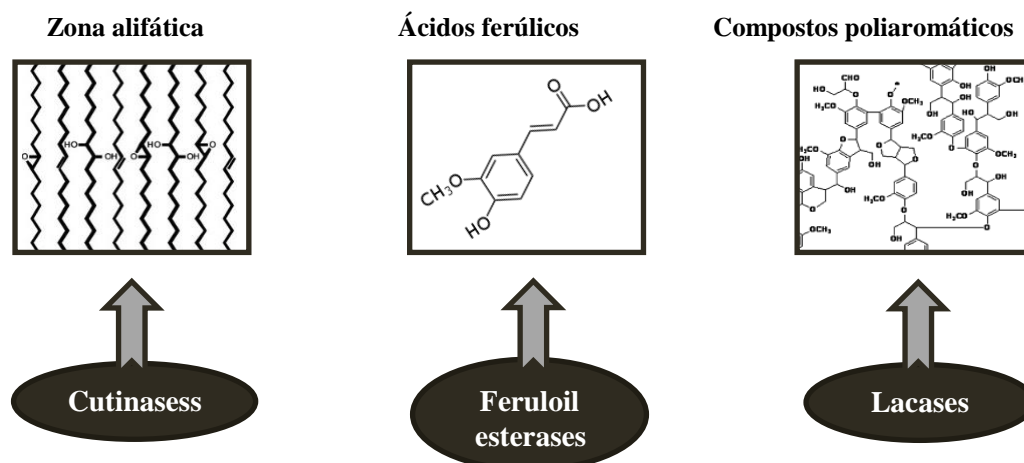


Figura 5 – Locais alvo de cada enzima (suberinase) na molécula de suberina.

2.1 Cutinases

As cutinases (EC 3.1.1.74) são poliesterases, uma subclasse de hidrolases éster carboxílicas (EC 3.1.1), que constituem o membro mais pequeno da superfamília α/β -hidrolase [18; 19]. São maioritariamente produzidas por fungos, seguindo-se as bactérias, e estão também presentes no pólen de plantas [16; 19].

Nos fungos patogénicos, estas enzimas estão envolvidas na hidrólise enzimática da barreira protetora de cutina da planta, no estágio inicial da infeção [16; 19]. Nos últimos anos, as cutinases foram descritas em vários géneros de microrganismos como: *Actinomyces*, *Alternaria*, *Aspergillus*, *Bacillus*, *Botrytis*, *Candida*, *Chromobacterium*, *Chrysonilia*, *Coprinopsis*, *Fibrobacter*, *Fusarium*, *Geotrichum*, *Humicola*, *Penicillium*, *Pseudomonas*, *Rhizomucor*, *Rhizopus*, *Streptomyces*, *Thermomonospora*, *Trametes*, *Trichoderma* ou *Venturia* [16; 17; 18; 19].

As cutinases têm a capacidade de hidrolizar uma variedade de ésteres sintéticos, domínios alifáticos de tecidos suberosos e mostram atividade tanto em triacilglicerídeos solúveis como em emulsificados de cadeia curta ou longa, com uma eficiência semelhante às lipases pancreáticas [19]. As sequências codificantes destas enzimas exibem 190-230 aminoácidos, com um peso molecular que pode variar entre 19 a 26 kDa para provenientes de fungos, e aproximadamente 30 kDa para derivadas de bactérias [16; 19].

Os centros ativos das cutinases são constituídos por resíduos de serina, histidina e ácido aspártico, originando uma tríade-catalítica, semelhante a várias lipases e proteases de serina (Figura 6) [19]. Contudo, diferem das lipases clássicas dado que não exibem ativação interfacial [18; 19]. Uma pequena sequência característica de cutinases em volta do sítio ativo de serina é Gli-Tir-Ser-Gln-Gli [16; 18].

É também importante referir que, apenas um pequeno conjunto de enzimas que modificam lípidos está disponível no mercado e, em geral, as lipases clássicas não exibem atividade enzimática perante polímeros como cutina e suberina, dado que são altamente ativas na interface entre o substrato (lípidos insolúveis) e água [16]. Deste modo, a nível industrial, as cutinases inserem-se num patamar de grande importância porque apresentam propriedades peculiares com atividades superiores às das lipases e, adicionalmente, boa estabilidade e atividade na presença de agentes oxidantes [19]. Podem ser aplicadas numa diversa variedade de setores comerciais: despolimerização de plásticos, síntese de ésteres,

agricultura, modificação de fibras de poliéster, detergentes, indústrias farmacêutica e alimentar, e ainda em lavagens para remoção de gorduras e algodão [16; 19].

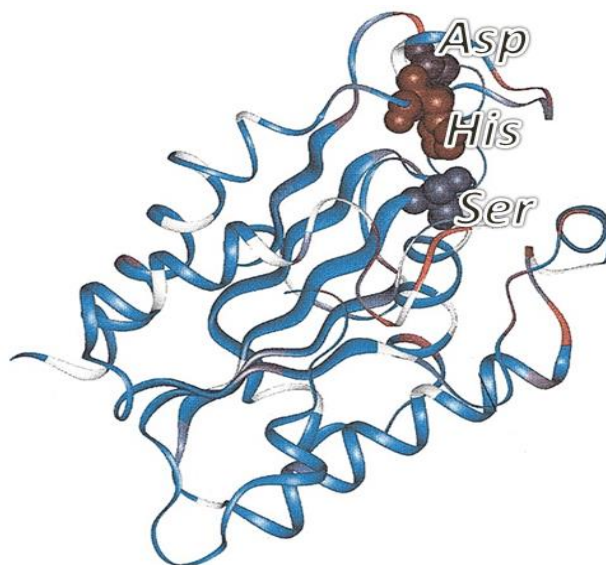


Figura 6 – Representação da estrutura de uma cutinase com a triade-catalítica exposta (Ser, His, Asp) adaptado [20].

2.2 Feruloil esterases

As feruloil esterases (FAEs, EC 3.1.1.73) são uma subclasse de hidrolases éster carboxílicas (EC 3.1.1), também conhecidas como cinamoil esterases, cinamoil éster hidrolases, feruloil/p-comariol esterases e ácido ferúlico/p-comárico esterases [21]. As FAEs desempenham um papel fundamental na degradação da estrutura complexa da parede celular vegetal, por hidrólise dos grupos ésteres de ácidos hidroxicinâmicos, essencialmente ácido ferúlico e comárico, envolvidos nas ligações entre hemiceluloses e entre a hemicelulose e a lenhina [22].

Durante a última década, várias FAEs produzidas por bactérias e fungos foram purificadas e caracterizadas, apresentando grandes variações em propriedades químicas como o ponto isoelétrico e peso molecular entre 27 a 210 kDa, e nas condições de reação como valores ótimos de pH e temperatura diferentes [22; 23]. As sequências que codificam estas enzimas diferem de 280 a 520 aminoácidos, e alguns gêneros de microrganismos que as produzem são: *Aspergillus*, *Butyrivibrio*, *Cellvibrio*, *Clostridium*, *Fusarium*, *Lactobacillus*, *Neocallimastix*, *Neurospora*, *Penicillium*, *Piromyces*, *Streptomyces* ou *Talaromyces* [21; 24].

A sua classificação é feita em quatro tipos diferentes (A, B, C e D), tendo por base a sequência de aminoácidos que codifica a enzima e a especificidade exibida nos seguintes substratos fenólicos: ácido ferúlico, ácidos diferúlicos, ácido comárico, ácido sinapínico e ácido caféico [22; 24]. Estas comparações indicam uma relação evolutiva entre FAEs, acetil-xilano esterases (EC 3.1.1.72) e algumas lipases [22; 23]. Quanto à estrutura cristalina, em geral, apresenta uma conformação α/β -hidrolase com uma tríade-catalítica formada por uma serina, uma histidina e um ácido aspártico, característico de lipases (Figura 7) [22].

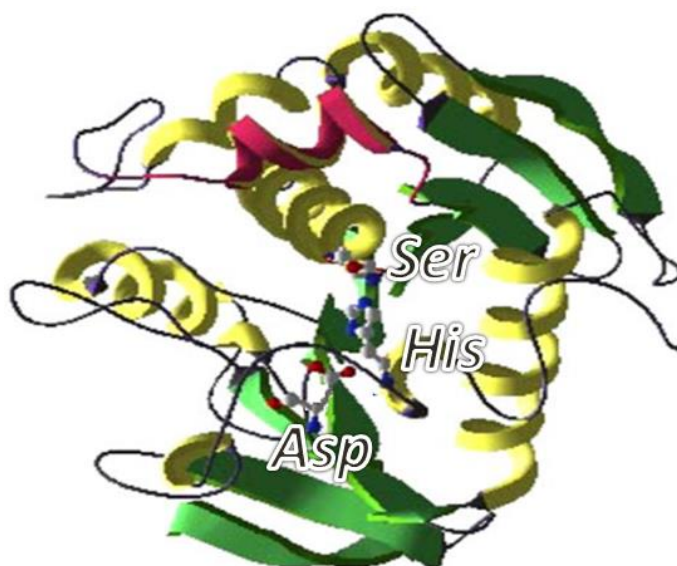


Figura 7 – Representação 3D de uma feruloil esterase e da sua tríade-catalítica (Ser, His, Asp), proveniente de *Aspergillus niger* adaptado [21].

As ligações cruzadas de ácidos ferúlicos, entre a rede intrínseca de componentes da parede celular, influenciam propriedades como extensibilidade, plasticidade e digestibilidade; o que, por conseguinte, aumenta a complexidade desta estrutura celular e limita o acesso aos polissacarídeos, dificultando a hidrólise enzimática [22]. Os ácidos ferúlicos obtidos pela hidrólise enzimática das FAEs são antioxidantes naturais, com potenciais aplicações nas indústrias farmacêutica e alimentar; além disso, os ácidos hidroxicinâmicos têm vindo a ganhar interesse como matéria-prima para posterior conversão em compostos de valor acrescentado como estirenos, polímeros, alquil benzenos, derivados de ácido vanílico, guaiacol, catecol e vanilina [24].

Estudos anteriores indicam que a hidrólise enzimática de FAEs melhoram as condições de uso de enzimas lenholíticas na bioconversão de material lenhocelulósico em açúcares fermentáveis, para produção de bioetanol [22]. Diferenças subtis de especificidade, muitas vezes desvalorizadas, podem ser importantes para determinar a sinergia ideal entre FAEs e outras hemicelulases, na eficiência de hidrólise das paredes celulares [22]. Como resultado, a descoberta de novas FAEs continua a ser uma área de investigação bastante ativa [22; 24].

2.3 Lacases

As lacases (p-difenol oxidase, EC 1.10.3.2) pertencem à família de proteínas multi-cobre oxidases e são produzidas por uma variedade de plantas, fungos, insetos, e também bactérias [25; 26]. São proteínas glicosiladas, geralmente contendo menos monossacarídeos (10-25%) em fungos do que em plantas [25]. Apresentam uma mobilidade em SDS-PAGE correspondente a um peso molecular entre 60 a 100 kDa, consoante o seu nível de glicosilação; e sequências entre 510 a 620 aminoácidos, sendo filogeneticamente muito próximas entre si [25].

De acordo com a especificidade de substrato, as lacases são classificadas como proteínas de cobre azul e entram, também, na descrição geral das polifenol oxidases, uma vez que são capazes de oxidar compostos aromáticos, utilizando oxigénio molecular como recetor final de eletrões [26]. Estas enzimas catalizam a oxidação de uma enorme quantidade de substratos orgânicos e inorgânicos, incluindo mono-, di-, e polifenóis, aminofenóis, metoxifenóis, aminas aromáticas e ascorbato, com a concomitante redução de oxigénio para água [27].

As lacases contêm quatro átomos de cobre, localizados em zonas ricas de histidina, e são classificados em três tipos (T1, T2 e T3), apresentando diferentes propriedades espectroscópicas e diferindo consoante o seu ambiente [28]. Os cobres T2 e T3 combinam-se para formar uma estrutura trinuclear que está envolvida na ligação com o oxigénio, e está coordenada por 8 resíduos de histidina (Figura 8) [26; 28]. O cobre T1 é composto por 2 resíduos de histidina e 1 resíduo de cisteína, e está relacionada com o processo de captura de eletrões [26; 28].

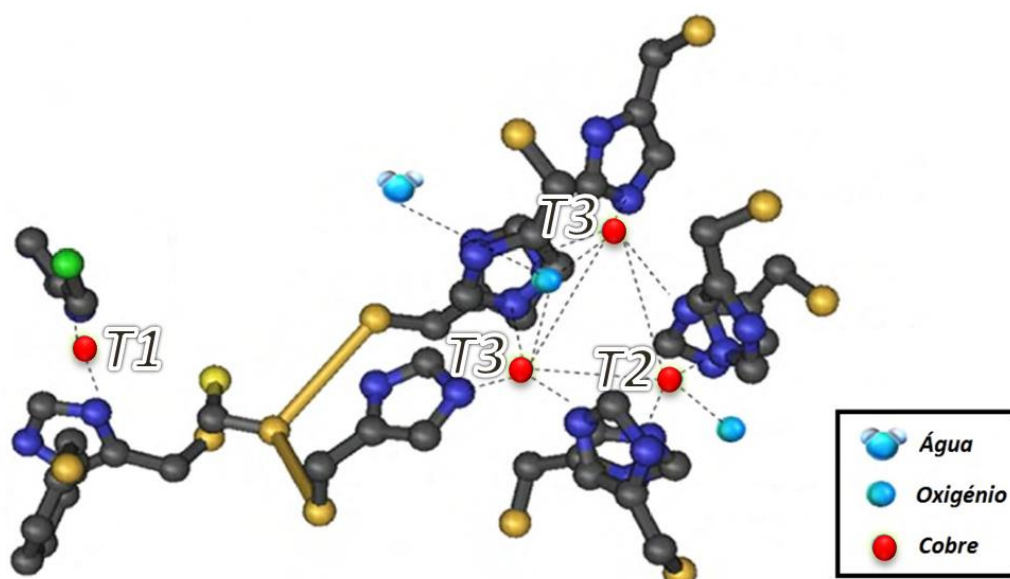


Figura 8 – Representação esquemática dos átomos de cobre de uma lacase e da sua estrutura trinuclear ^{adaptado [29]}.

A grande maioria das lacases descritas são proteínas extracelulares, porém foram já detetadas lacases intracelulares em vários fungos e insetos [25]. Estas enzimas têm sido isoladas e caracterizadas, principalmente em fungos, mais propriamente, ascomicetos, deuteromicetos e basidiomicetos [30]. A sua função difere de organismo para organismo, por exemplo, nos fungos é responsável pela remoção de fenóis tóxicos do meio onde crescem sob condições naturais; no caso das plantas estão envolvidas em processos de síntese, como a polimerização de lenhina [31].

A degradação microbiana de lenhina tem sido bastante estudada em fungos de podridão branca (basidiomicetos), mas em contraste, pouco é conhecido sobre a degradação desta estrutura fenólica por parte das bactérias [32]. Estudos recentes sugerem que bactérias presentes em vários solos são capazes de degradar lenhina [33] e para além disso, progressos na análise de genomas indicam que estas enzimas estão amplamente difundidas em bactérias [26; 34; 35]. Alguns géneros de bactérias descritas que reportam a atividade desta enzima são: *Arthrobacter*, *Azospirillum*, *Burkholderia*, *Bacillus*, *Escherichia*, *Marinomonas*, *Nocardia*, *Pseudomonas*, *Rahnella*, *Rhodococcus*, *Sphingomonas*, *Streptomyces*, *Xanthomonas* ou *Yersinia* [25; 32; 34; 35].

Atualmente, as lacases são enzimas extremamente importantes a nível industrial, uma vez que podem ser usadas para um diverso número de aplicações como deslenhificação de material lenhocelulósico, detoxificação e descontaminação de desperdícios, descoloração

de corantes têxteis, na indústria alimentar, em aplicações medicinais e cuidados pessoais, em biosensores e testes de diagnóstico [25; 30].

Hoje em dia, apenas as lacases provenientes de fungos são utilizadas em processos biotecnológicos [32]. Contudo, uma expressão eficiente de lacases recombinantes, usualmente requerida para aplicações tecnológicas, é bastante mais difícil de obter em fungos do que em bactérias [36]. A formação de pontes dissulfídicas e glicosilação de lacases, o desconhecimento de informação de sequências e as estruturas exão-intrão de genes eucarióticos, são frequentemente obstrutivos [36].

Em suma, as bactérias apresentam-se como um emergente potencial da bioconversão de lenhina em compostos químicos aromáticos, estando já a ultrapassar desvantagens na instabilidade e aplicações no processo, verificadas em fungos [35]. Foram já identificadas lacases bacterianas muito ativas e muito mais estáveis a temperaturas e valores de pH altos, o que leva à constante e incessante procura de novas lacases com propriedades melhoradas [26; 33].

3. Fontes naturais de suberinases – *Crematogaster scutellaris*

A omnipresença das formigas depende não só da sua alta amplitude ecológica, da sua organização social fortemente desenvolvida e do seu comportamento complexo, mas também da sua espantosa habilidade em explorar uma vasta gama de recursos alimentares [37]. As formigas são portanto, bons indicadores biológicos que permitem refletir a qualidade/impacto ambiental e funcionamento dos ecossistemas [37].

O papel desempenhado pelas formigas nos sistemas florestais, mais concretamente, as complexas interações entre formigas e plantas têm sido alvo de estudo de muitos autores [38]. Em Portugal, foi já analisada a diversidade de formigas presentes em três tipos de ecossistemas florestais: eucaliptal, pinhal e montado [38]. Dando particular importância aos montados, num total de 16 espécies de formigas identificadas, a mais comum é, sem dúvida, *Crematogaster scutellaris* (Hymenoptera, Formicidae), uma das principais pragas do sobreiro que afeta pranchas de cortiça [38; 39].

3.1 *Crematogaster scutellaris* e a sua atividade nos sobreiros

O género *Crematogaster* é típico da região do Mediterrâneo Ocidental [40; 41]. A espécie *Crematogaster scutellaris* prevalece sobretudo na Península Ibérica, onde é mais ativa desde a Primavera até ao Outono, podendo ocasionalmente ser observada no Inverno, se houver longos períodos com temperaturas elevadas e sem chuva [39; 41].

A maioria das formigas são omnívoras e o género *Crematogaster* é caracterizado por ter um regime essencialmente insetívoro e por dependerem de substâncias ricas em açúcares (meladas) segregadas por afídeos [40; 41]. Desta forma, estas formigas procuram com frequência diversos hemípteros, colocando-os em pontos estratégicos de plantas nutritivas, sendo também capazes de armazená-los nos seus ninhos durante o Inverno [41].

C. scutellaris apresenta um corpo bicolor, de cabeça avermelhada com a parte posterior do corpo escurecida (Figura 9) [40]. Em geral, esta formiga vive na base ou no interior da casca de árvores vivas, onde a madeira está parcialmente danificada, queimada ou degradada por fungos; outras vezes, nidifica em troncos caídos [41].



Figura 9 – Imagem correspondente a uma obreira da espécie *Crematogaster scutellaris* ^[42].

Esta espécie apresenta uma grande flexibilidade no seu comportamento nidificador, usufruindo de sobreiros como habitat preferencial e desenvolvendo-se sobretudo na cortiça virgem, que se obtém da primeira extração (Figura 10-A) [39; 41]. As colónias constroem numerosas galerias e camaras, à volta do tronco, em quase toda a espessura da cortiça [39; 41]. Por outro lado, representam uma relação de mutualismo perante os sobreiros, protegendo-os contra pulgões e cochonilhas [41].

A cortiça de reprodução ou amadia, que se desenvolve após o segundo descortiçamento e se destina primordialmente ao fabrico das rolhas, pode ser também colonizada, provocando uma depreciação de cerca de 25% na cortiça, dependendo do quão graves forem as marcas deixadas pelas formigas (Figura 10-B) [41]. Adicionalmente, atacam as pilhas de cortiça armazenadas que permanecem muito tempo sem tratamento, causando danos que podem ser significativos [41].

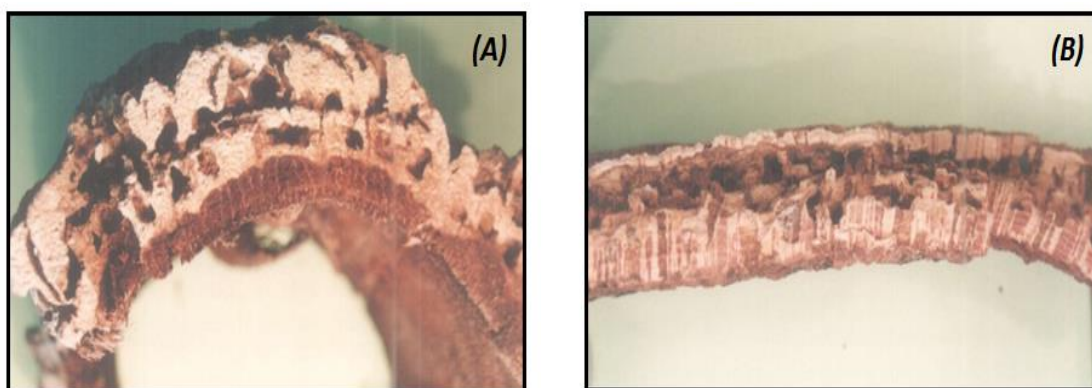


Figura 10 – Danos causados pelas formigas na cortiça virgem (A) e na cortiça de reprodução (B) adaptado [41].

De uma forma geral, pensa-se que *C. scutellaris* constrói os seus ninhos a partir de um substrato (cortiça), que ela própria consegue transformar em aglomerados de serrim, utilizando a saliva [41]. Apesar de os ninhos no interior da cortiça não serem profundos e não chegarem a atingir a madeira, são bastante extensos [41]. As formigas aplicam também o serrim para recobrir e separar cavidades demasiado grandes, consolidar paredes, obstruir orifícios ou galerias; mas não é aplicado no exterior do ninho porque é facilmente degradado por agentes atmosféricos [41].

3.2 Sistemas enzimáticos de degradação de material lenhocelulósico em insetos

Aguns insetos alimentam-se de madeira e têm a aptidão de contornar a barreira de lenhina e auferir do acesso a hidratos de carbono, como fonte de energia [43]. Um exemplo, bastante estudado, é o caso das térmitas por apresentarem eficiência elevada na digestão de material lenhocelulósico, preferencialmente madeira morta ou pré-degradada [44]. Outros insetos destacam-se por se desenvolverem no interior da madeira de várias

árvores folhosas, como é o caso dos escaravelhos, por intermédio de uma relação exo-simbionte com um fungo degradador de madeira [43; 45].

As térmitas digerem material lenhocelulósico através de enzimas digestivas endógenas e exógenas produzidas por simbioses no intestino [46]. Análises recentes de transcriptomas de simbioses e hospedeiros (térmitas) revelaram genes codificantes de celulases, hemicelulases, enzimas degradadoras de lenhina, e de outras enzimas digestivas relevantes [46]. No genoma da térmita foram também encontrados genes que codificam enzimas envolvidas na hidrólise de lenhina, suportando a hipótese de que estas enzimas não são apenas derivadas de simbioses do intestino [44; 46].

Apesar das térmitas secretarem as suas próprias enzimas digestivas, a digestão de alimentos recalcitrantes, como a madeira, depende largamente da diversidade dos seus microrganismos presentes no intestino, compreendendo protistas, bactérias e archaea [47; 48; 49]. Abordagens moleculares de técnicas independentes de cultivo indicam que a maioria das comunidades bacterianas, presentes no intestino de insetos, consiste em várias espécies que ainda não estão caracterizadas, portanto desconhecidas para os microbiólogos [48; 49; 50]. O microbioma intestinal simbiótico de insetos é vital para a utilização e especialização em plantas de diferentes qualidades nutritivas; para além disso, é considerado um recurso importante para a descoberta de biocatalisadores e microrganismos com inúmeras aplicações biotecnológicas [51].

Paralelamente, as formigas apresentam-se como alvo de estudo para detecção de novas enzimas, por também explorarem uma vasta variedade de materiais lenhocelulósicos como madeira e cortiça [41; 52]. O seu sistema digestivo representa todo o comprimento do corpo como um tubo, desde a boca até ao ânus, com uma tendência evolutiva no gáster (Figura 11) [53]. Os resíduos sólidos e/ou semi-sólidos ao entrarem na boca (bolsa infrabucal) entram em contato com os vários compostos provenientes das glândulas [54]. No “crop” são retidos os resíduos alimentares, sendo posteriormente passados lentamente, em quantidades controladas, para o estômago (órgão principal digestivo), através de um género de válvula bombeadora – proventrículo [55]. Posteriormente ao estômago, encontra-se o intestino que consiste num tubo circular e fino [53]. Por sua vez, os túbulos de Malpighian situam-se próximo da junção entre o estômago e o intestino.

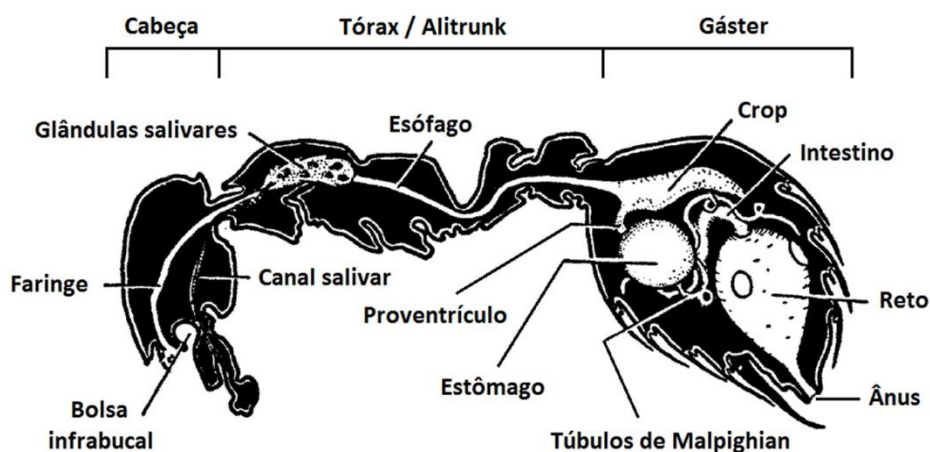


Figura 11 – Representação dos vários componentes do sistema digestivo da formiga ^{adaptado [56]}.

A exploração e o conhecimento destes sistemas enzimáticos têm grande potencial para identificação de enzimas com aplicação no pré-tratamento eficiente de material lenhocelulósico, uma vez que a despolimerização de hemicelulose e lenhina ainda não está bem desenvolvida a nível industrial [44].

4. Estudos de comunidades microbianas

4.1 Metagenómica

As abordagens metagenómicas têm revolucionado a análise de comunidades microbianas, uma vez que permitem uma visão mais detalhada das funções e interações que ocorrem entre microrganismos e o ambiente onde se inserem [57]. A metagenómica define-se como a análise de DNA genómico de uma comunidade microbiana completa (metagenoma), distinguindo-se da genómica, que apenas estuda o DNA genómico de um organismo individual ou célula [57].

O termo foi inicialmente descrito em 1998 através de clonagem aleatória de DNA ambiental, mais propriamente, num estudo de microrganismos presentes em solos [58]. Ademais, as abordagens metagenómicas são independentes de cultivo, permitindo o estudo de DNA genómico de comunidades, o que torna esta técnica bastante ampla [57].

A metagenómica aplica diversas tecnologias genómicas e ferramentas bioinformáticas, dividindo-se em duas áreas: pesquisa de apenas um gene e estudo de todos os genes

ambientais [57; 59]. O primeiro caso pode ser visto como um estudo metagenômico direto e focado [57]. Os genes alvo são amplificados por reação de amplificação em cadeia (PCR). Posteriormente os produtos são sequenciados, fornecendo a identificação de todas as versões do mesmo gene presentes nos diferentes organismos de uma dada comunidade [59].

Na segunda área é estudada a totalidade do DNA isolado de uma amostra ambiental e em seguida sequenciado, resultando num perfil de todos os genes presentes na comunidade. Os genes são anotados, são identificadas as proteínas que podem ser sintetizadas pelo metagenoma e inferida a sua relevância no ambiente em estudo [57]. Do mesmo modo, pode ser inferido o proteoma de uma comunidade [60] e por último o seu estado metabólico por análise das potenciais proteínas em consumir ou produzir metabolitos (Figura 12) [61].

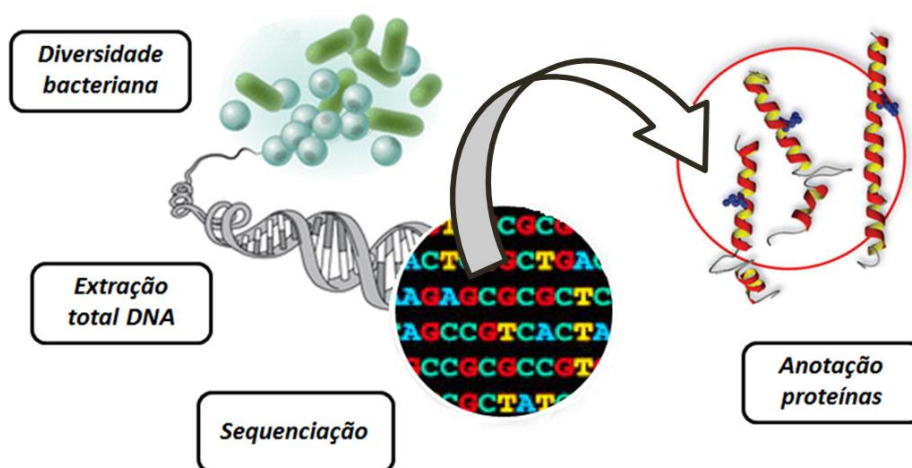


Figura 12 – Metodologia utilizada na abordagem metagenômica que estuda todos os genes presentes numa amostra ambiental ^{adaptado [57]}.

4.2 Métodos de determinação da estrutura e composição de comunidades microbianas

Na natureza, as bactérias encontram-se sobretudo em associação com outras de diferentes espécies, estabelecendo deste modo as comunidades bacterianas [62]. Estas comunidades estão presentes em todo o tipo de ambientes desde solo, água, ar, plantas e até mesmo animais [62].

De forma a superar várias limitações associadas aos métodos de isolamento e cultivo, dado que pelo menos 99% dos microrganismos presentes na natureza são incultiváveis [63], os avanços surpreendentes na biologia molecular permitiram o desenvolvimento de abordagens independentes de cultura, com o intuito de descrever, caracterizar e comparar comunidades bacterianas [64]. Este tipo de abordagens abrange métodos como lise celular, extração de ácidos nucleicos, amplificação por PCR, clonagem e sequenciação de DNA [65].

O estudo das comunidades microbianas é feito com base em dois tipos de análise [65]:

- ✓ Completa – ferramenta molecular que visa a análise total da informação genética, a partir do DNA extraído;
- ✓ Parcial – ferramenta molecular que normalmente explora frações da informação, focando-se em sequências-alvo de genomas, amplificadas por PCR.

Na análise completa de DNA, algumas das técnicas disponíveis são hibridização do DNA (DDH) [66], cinética de reassociação [67] e fracionamento de DNA através do conteúdo de bases G+C por gradiente de densidade [68].

Na análise parcial do DNA, as sequências alvo que se utilizam frequentemente são os genes do operão ribossomal, mais propriamente os genes do rRNA 16S para bactérias [69] e a região intergénica do RNA ribossomal (ITS) para fungos [70]. Estes métodos integram a clonagem de fragmentos por PCR seguido da análise por restrição e/ou sequenciação, o que permite o conhecimento da diversidade da comunidade não só em números de espécies diferentes, mas também na abundância relativa de espécies [65].

Adicionalmente, outra alternativa de análise parcial é a impressão do perfil genético que proporciona uma imagem global da estrutura genética da comunidade bacteriana, porém não permite uma identificação taxonómica direta [65]. Este tipo de abordagem molecular baseia-se na separação das sequências amplificadas, através da migração diferencial em gel de agarose ou acrilamida, e/ou por sequenciação [65]. A separação é realizada conforme o polimorfismo do tamanho do fragmento de restrição (T-RFLP) [71], o espaço intergénico ribossomal (RISA) [72], o DNA polimórfico amplificado aleatoriamente (RAPD) [73], o DNA ribossomal amplificado (ARDRA) [74], ou consoante o polimorfismo da sequência, como é o caso de eletroforese em gel com gradiente de desnaturação (DGGE) ou de temperatura (TGGE) [75].

Na última década, a metagenómica tem vindo a ganhar particular importância devido ao desenvolvimento das tecnologias de sequenciação de última geração [76; 77]. Dentro

destas, a sequenciação 454 dos genes do rRNA 16S destaca-se por apresentar várias vantagens em relação a outros métodos, substituindo o método de sequenciação de Sanger [78]. Este método é apropriado e bastante eficaz na identificação de microrganismos, mais propriamente, na sequenciação da região hipervariável do rDNA 16S [77].

4.2.1 Pirosequenciação dos genes do rRNA 16S

Atualmente, a sequenciação de DNA é, sem dúvida, uma das mais importantes técnicas para o estudo de sistemas biológicos [79]. O método convencional de sequenciação foi desenvolvido por Sanger *et al.* em 1977 [80]. Desde então têm surgido vários progressos com o objetivo de adquirir características superiores como velocidade, sensibilidade e rendimento, dado que são de extrema importância para o estudo de diversos genomas de bactérias, archaea e de eucariotas [79].

A par do desenvolvimento tecnológico, a constante procura de técnicas robustas e mais eficientes levou ao aparecimento de métodos de sequenciação múltipla paralela, entre elas a sequenciação 454 da Roche [81]. Esta técnica é largamente utilizada na sequenciação de genomas, transcriptomas, amplicões ou metagenomas, e apresenta diversas vantagens na precisão, flexibilidade, processamento paralelo, assim como na automatização [79].

Um estudo recente comparou a sequenciação 454 do rDNA 16S com outros métodos tradicionais como Sanger e cultivo de bactérias, na exploração da diversidade bacteriana presente em diferentes tecidos de formigas [82]. Os autores concluíram que mesmo depois de ser aplicada uma filtragem de qualidade rigorosa, a pirosequenciação obteve 445 unidades operacionais taxonómicas (OTUs) não detetadas pelos restantes métodos, realçando uma diversidade bacteriana significativa em tecidos intestinais comparativamente a outros tecidos, o que indica uma profunda coevolução entre as formigas e microrganismos simbiotes de função nutritiva [82]. Aliás, inferem a sequenciação 454 como uma ferramenta de custo eficiente para a monitorização da diversidade bacteriana presente nas formigas [82].

Este tipo de abordagem metagenómica consiste na identificação de microrganismos presentes numa comunidade complexa, recorrendo à exploração de alvos universais e conservados, como é o caso dos genes do RNA ribossomal [77]. Através da amplificação de regiões selecionadas dentro do rDNA 16S, podem ser identificadas essencialmente

bactérias e archaeas [77]. A amplificação por PCR é feita com primers em regiões conservadas, amplificando regiões hipervariáveis que irão facilitar a identificação de géneros e espécies [77]. No caso das bactérias, o gene do rRNA 16S consiste em sequências conservadas (C) intercaladas com 9 sequências hipervariáveis (V) (Figura 13) [77].

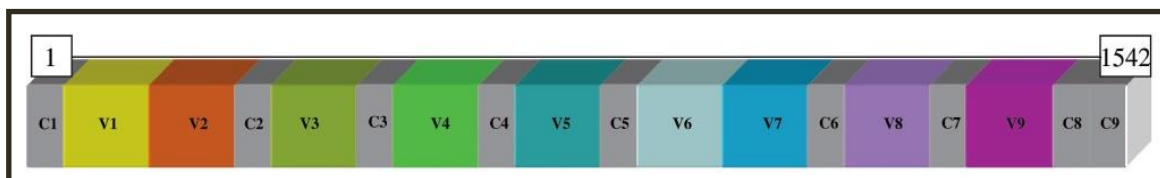


Figura 13 – Representação das regiões conservadas (C) e hipervariáveis (V) do gene rRNA 16S, correspondente a um total de 1542 pares de bases (pb) ^{adaptado [77]}.

Este método necessita de uma biblioteca de amplicões, que consiste numa mistura complexa de sequências amplificadas pelos primers de fusão especiais [83; 84]. Os amplicões produzidos deverão compreender entre 200-600 pb, sendo fundamental o conhecimento detalhado dos locais da sequência alvo reconhecidos pelos primers [85]. Na extremidade 5' do primer “forward” está presente um adaptador do tipo A, para mais tarde ser combinado com outros componentes da sequenciação, enquanto que, o primer “reverse” integra um adaptador do tipo B [85]. Na extremidade 3' de cada primer são incluídos os primers específicos para as regiões consenso a amplificar, geralmente com 20 a 25 pb [85]. Os primers ainda possuem uma sequência diferenciadora entre os adaptadores A ou adaptadores B e a sequência específica, os marcadores, de forma a reconhecer amplicões derivados de diferentes amostras [83; 85]. Após amplificação, cada amplicão gerado fica ladeado por um adaptador A na extremidade 5' e um adaptador B na extremidade 3'.

Posteriormente à construção da biblioteca de amplicões, estes são imobilizados em “beads” que capturam DNA, numa proporção que garanta que a cada “bead” se ligue apenas uma molécula de DNA. Nesta fase, o adaptador B que foi ligado à extremidade 3' de cada molécula de DNA molde durante a preparação da biblioteca, hibridiza com oligonucleótidos complementares que estão ligados às “beads”. Desta forma, em cada “bead” existe apenas um tipo de amplicão (clone), que será amplificado de forma clonal no passo seguinte, o PCR de emulsão [84].

O PCR é preparado por emulsificação, o que envolve uma agitação mecânica de uma mistura lípido-aquosa sob condições controladas [84]. Nesta etapa, os produtos são micelas, designadas por micro-reatores, de 50 a 100 µm de diâmetro (fase aquosa), em que cada micro-reator contém todos os componentes da mistura de amplificação e apenas uma “bead” [84]. As “beads” são submetidas a PCR clonal para amplificação de cada molécula de DNA molde, também designado como PCR de emulsão (emPCR) [84]. O número de cópias de DNA molde, presente em cada “bead” no final da amplificação, varia entre 10 a 50×10^6 [84].

Finalizada a amplificação, a emulsão é quebrada quimicamente e as “beads” que transportam as moléculas de DNA amplificadas em cadeia dupla são recuperadas e lavadas. Nesta fase, decorre o enriquecimento de todas as “beads” que contêm a biblioteca de DNA amplificado, ou seja, são removidas as “beads” que não contêm DNA amplificado ou que contêm fragmentos de DNA pouco ou mal amplificados. Este processo envolve a hidrização entre o adaptador de cada molde amplificado e primers enriquecidos com biotina, que por sua vez se ligam a “beads” magnéticas revestidas com estreptavidina [84]. Seguidamente, as “beads” magnéticas são separadas das “beads” que contêm o DNA amplificado através da desnaturação dos produtos da amplificação, removendo-se os primers enriquecidos em biotina e criando-se uma população de “beads” que carregam moléculas de cadeia simples – este será o material introduzido no sequenciador [84]. As “beads” são depositadas numa placa com milhões de poços, onde em cada poço será apenas sequenciada um tipo de molécula (Figura 14) [79].

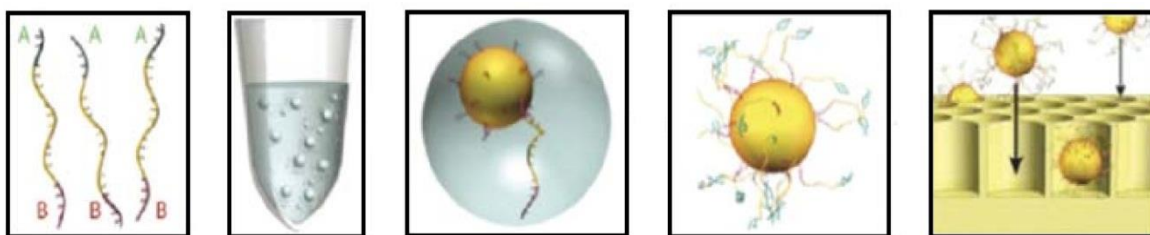


Figura 14 – Representação das várias etapas da pirosequenciação desde a separação da biblioteca de amplicões (esquerda) até à entrada das “beads” na placa de sequenciação (direita) adaptado [79].

A pirosequenciação baseia-se na detecção de pirofosfato (PPi) libertado durante a síntese de DNA (Figura 15) [79]. Numa série de reacções enzimáticas, a luz visível produzida é proporcional ao número de nucleótidos incorporados. As várias reacções começam com a libertação de PPi inorgânico à medida que a DNA polimerase incorpora os ácidos nucleicos na cadeia simples [86]. Na presença de adenosina 5'-fosfossulfato (APS), os PPi são subsequentemente convertidos a ATP pela ATP sulfúrilase, que por sua vez fornece energia à luciferase para oxidar luciferina e produzir luz [79]. A luz é detetada por um dispositivo de carga acoplado, originando uma série de picos – o pirograma. Uma vez que se conhecem os nucleótidos adicionados e a sua ordem, cada sequência de DNA molde é determinada em apenas um pirograma [87].

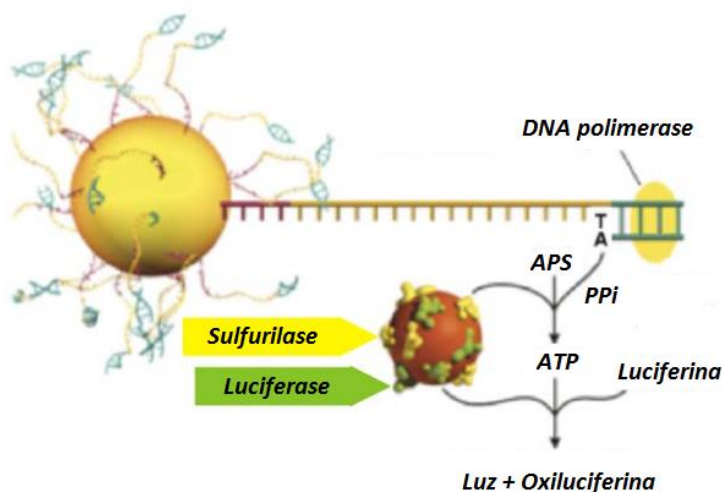


Figura 15 – Série de reacções enzimáticas que produzem luz durante a pirosequenciação ^{adaptado [79]}.

4.3 Métodos de determinação de genes alvo codificantes de enzimas

As amostras de DNA ambiental são extremamente ricas em genes codificantes de enzimas relevantes para processos biotecnológicos [88]. A pesquisa de genes alvo proporciona uma visão extensa de padrões ecológicos, evolutivos e funcionais de genes importantes, presentes em microrganismos de vários ecossistemas [89]. Esta abordagem permite a análise e a aquisição de uma diversidade de genes presentes em metagenomas [89; 90].

Apesar dos métodos para interpretar dados metagenómicos ainda permanecerem numa fase inicial de desenvolvimento, esta área de investigação é bastante mais efetiva para o

conhecimento do conteúdo e composição de genes relevantes em comunidades bacterianas [89]. Estratégias baseadas na função e sequência de genes metagenômicos foram já estabelecidas para isolar, clonar e expressar novos genes [88], o que permite uma maior cobertura e uma redundância extensiva de sequências de genes alvo, revelando zonas específicas de genomas diretamente ligadas à função ecológica, mesmo com uma abundância muito baixa destes genes no metagenoma [89].

Para se aceder à enorme diversidade de genes presentes em metagenomas, existem principalmente três tipos de abordagens [88]:

- ✓ Clonagem de fragmentos de DNA ambiental e teste seletivo de atividade exibida por esses fragmentos em vetores recombinantes;
- ✓ Clonagem de fragmentos de DNA e montagem das sequências obtidas em genomas parciais;
- ✓ Amplificação por PCR, através da utilização de primers “consenso” degenerados que são específicos para uma classe de enzimas, seguido por “genome walking”.

O teste seletivo de atividade, exibida por fragmentos de DNA ambiental em vetores recombinantes, não depende do conhecimento prévio de qualquer sequência, o que torna este método capaz de detetar genes codificantes de novas enzimas, com estruturas e funções desconhecidas [88]. Todavia, é necessário testar centenas de milhares de fragmentos de DNA recombinantes, e além disso, as atividades exibidas podem não representar as características enzimáticas nativas [88].

Como alternativa, o método baseado na amplificação por PCR juntamente com “genome walking” é independente da atividade enzimática e conveniente para uma rápida expansão de bibliotecas de várias enzimas [88]. Apesar destas enzimas, correspondentes aos genes adquiridos por PCR, se assemelharem com as sequências previamente conhecidas da respetiva classe de enzima (zonas “consenso” de enzimas), esta metodologia possibilita a obtenção em larga escala de novas enzimas funcionais com diferentes características e propriedades, sem necessitar de uma extensa quantidade de testes seletivos de atividade [88; 90; 91]. Paralelamente, várias enzimas com importância biotecnológica foram já obtidas e identificadas por PCR, utilizando primers degenerados em sequências conhecidas de amostras metagenômicas como por exemplo: celulasas [90], lacases [33], lipases [92], monoxigenases do citocromo P450 [93] e quitinases [94].

4.3.1 Pirosequenciação dos genes amplificados por PCR através de primers “consenso”

Esta abordagem envolve primers desenhados em zonas “consenso” de enzimas permitindo a detecção de novos biocatalisadores a partir de microrganismos desconhecidos [88]. Adicionalmente, esta abordagem possibilita o conhecimento da diversidade de genes diretamente responsáveis por funções importantes em ecossistemas, como ciclos biogeoquímicos, biodegradação, patogénese, resistência a antibióticos e sinalização celular [91].

Em geral, esta abordagem consiste nas seguintes etapas [88]: alinhamento múltiplo de sequências de genes que codificam membros de uma classe específica de enzimas; construção de primers “consenso” degenerados; extração total de DNA microbiano ambiental; amplificação por PCR dos genes alvo através de primers “consenso”, utilizando DNA ambiental como molde; e sequenciação dos amplicões gerados pelos primers “consenso”.

Várias enzimas que pertencem à mesma classe específica contêm regiões conservadas (“consenso”), podendo ser facilmente visualizadas através de alinhamentos múltiplos de sequências [88; 91]. Os primers “consenso” podem ainda incluir bases degeneradas, ou seja, para uma localização específica de um nucleótido podem existir duas ou mais bases nos genes de diferentes organismos, o que permite cobrir aminoácidos extremamente conservados em proteínas [91]. Além disso, os primers deverão gerar amplicões entre 200 a 600 pb, requeridos na emulsão de PCR da pirosequenciação 454 [85]. Deste modo, os primers “consenso” ou conjunto de primers poderão cobrir genes correspondentes a uma família de proteínas [88; 90; 91]. Por fim, aos amplicões gerados são incluídos adaptadores, os quais permitem a sequenciação de ambos os lados, possibilitando cobrir fragmentos de tamanho superiores aos recomendados na pirosequenciação.

Em suma, este conjunto de técnicas metagenómicas aliado à pirosequenciação e análise de dados por diversas ferramentas bioinformáticas permitem a obtenção de uma grande diversidade de genes semelhantes (bibliotecas), que codificam enzimas com diferentes especificidades de substratos, atividades específicas, enantiopreferências, enantioseletividades, regioseletividades, termoestabilidades, estabilidades em solventes e

pH, e inibições de substratos e produtos; acelerando desta forma a descoberta de enzimas, o que poderá satisfazer as necessidades de diversas indústrias (Figura 16) [90].

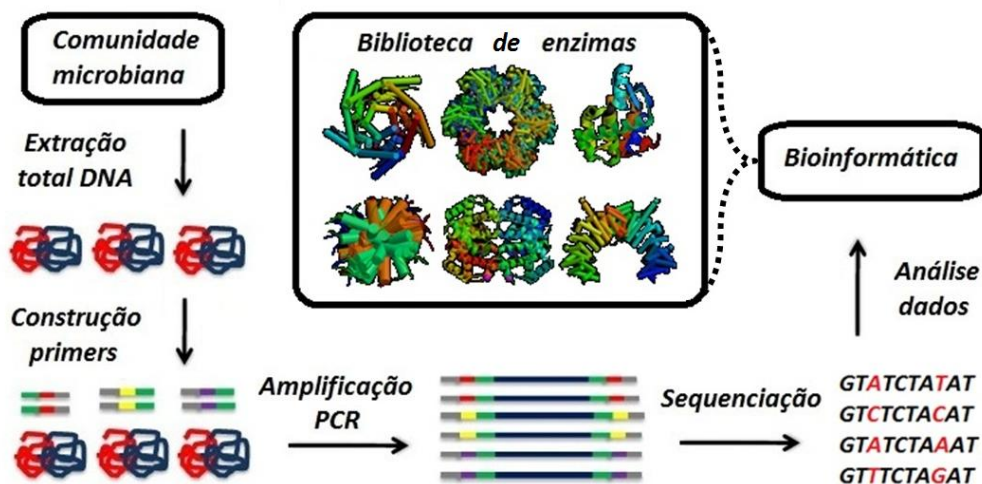


Figura 16 – Metodologia utilizada na abordagem metagenômica de pesquisa de genes alvo para criação de bibliotecas de enzimas adaptado [95].

4.4 Ferramentas bioinformáticas

A análise de enormes quantidades de dados produzidos pelos novos métodos de sequenciação, exige aplicações eficientes e flexíveis de bioinformática que tanto se ajustem às necessidades dos utilizadores, como às características dos dados de sequenciação.

Várias ferramentas bioinformáticas estão disponíveis para o processamento e agrupamento de sequências geradas pela sequenciação 454 da Roche. No caso do estudo da composição das comunidades microbianas, é utilizada uma “pipeline” de classificação automática onde, inicialmente, são separadas as sequências de diferentes amostras, através da sequência marcadora que incorporam, e são aplicados diversos parâmetros de qualidade com o objetivo de eliminar sequências de baixa qualidade [85]. Seguidamente, as sequências são agregadas em Unidades Taxonômicas Operacionais (OTUs), como é o caso dos genes do rRNA 16S, e é atribuída a taxonomia a cada uma destas OTUs, recorrendo a sequências de referência depositadas em bases de dados. Algumas ferramentas bioinformáticas que permitem a atribuição taxonômica são por exemplo RDP [96], CLOTU [97] e MG-RAST [98]. Por fim, é feita uma estimativa da riqueza e diversidade microbiana através de diferentes métodos estatísticos, o que permite uma análise simples e rápida.

De modo idêntico, na sequenciação de amplicões, como por exemplo os fragmentos de DNA gerados pela amplificação dos primers “consenso” degenerados, é aplicada uma filtragem de qualidade rigorosa nas sequências obtidas. No entanto, as sequências em vez de serem agrupadas em OTUs, são “assembladas” em contigs, e posteriormente analisadas e validadas com base na informação depositada em bases de dados públicas. Diversas ferramentas bioinformáticas estão disponíveis para classificar sequências, e baseiam-se na similaridade de sequências, na procura de padrões, e nas assinaturas de proteínas para identificar domínios e famílias [99]. Dentro destas, o InterPro destaca-se por ser um recurso de documentação que abrange a maioria de assinaturas de proteínas, fornecendo uma ferramenta valiosa para anotações de proteínas [99]. As sequências de proteínas são pesquisadas através do software InterProScan para identificar assinaturas em bases de dados como: Pfam, PROSITE, PRINTS, PRODOM, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, Gene3D e PANTHER [99]. Adicionalmente, existem outros programas online que permitem a visualização de alinhamentos múltiplos de sequências como Clustal [100] e MUSCLE (Multiple Sequence Comparison by Log-Expectation) [101]; e a construção de primers como iCODEHOP [102] e primers4clades [103].

5. Objetivos do trabalho e metodologias a utilizar

Neste projeto de investigação, pretende-se estudar o microbioma intestinal da formiga, espécie *Crematogaster scutellaris*, com o objetivo de encontrar genes bacterianos que codifiquem suberinases, ou seja, enzimas que consigam degradar a suberina presente nas paredes celulares da cortiça de sobreiros (*Quercus suber L.*). Uma vez que estas formigas nidificam na cortiça, acredita-se que têm a capacidade de degradar a suberina e, como tal, é provável que elas possuam suberinases nas secreções salivares e no microbioma intestinal.

Tendo em conta o objetivo descrito anteriormente, vai ser estudado o microbioma intestinal da formiga, através de uma abordagem metagenómica, o que permitirá encontrar genes codificantes de novas enzimas bacterianas com atividade em suberina. Para tal, o DNA da comunidade microbiana presente no intestino e estômago das formigas (metagenoma) será extraído, seguidamente, as sequências correspondentes aos genes do rRNA 16S serão amplificadas por PCR e sequenciadas, através do sequenciador GS FLX Titanium da Roche (Pirosequenciação 454). Esta análise metagenómica fornecer-nos-á um melhor conhecimento da comunidade bacteriana em termos de abundância e classificação taxonómica, permitindo identificar as espécies ou géneros mais relevantes.

Especificamente, a principal meta será encontrar enzimas alvo, através de amplificação dirigida de enzimas a partir do DNA isolado do gáster da formiga. A fim de avaliar e explorar os microrganismos simbioses para aplicações biotecnológicas, proceder-se-á à construção de primers “consenso” degenerados que reconheçam regiões conservadas de cutinases, feruloil esterases e lacases, conforme as espécies ou géneros identificados pela pirosequenciação dos genes do rRNA 16S recorrendo a bases de dados e ferramentas bioinformáticas. As regiões amplificadas serão pirosequenciadas e comparadas com as enzimas presentes em bases de dados para validar a sua função. Este tipo de abordagem metagenómica permitirá a identificação de uma vasta diversidade de genes alvo (biblioteca de amplicões), tanto a nível taxonómico como funcional, através da “pipeline” de análise MG-RAST.

II - MÉTODOS

1. Recolha das formigas – *Crematogaster scutellaris*

As formigas, espécie *Crematogaster scutellaris*, foram recolhidas do interior e exterior da cortiça de um sobreiro, *Quercus suber* L., existente no Biocant Park, Parque Tecnológico de Cantanhede, Portugal (40°21'38,10''N/08°36'24,85''W) (Figura 17-A). Para tal, foi utilizado um “Pooter” (Figura 17-B) aparelho que consiste num recipiente selado constituído por dois tubos. As formigas entram para o recipiente pelo tubo azul, ao mesmo tempo que são sugadas pelo tubo transparente, que está vedado por um filtro em baixo.

Seguidamente, as formigas foram transferidas para um tubo Falcon de 50 mL com etanol a 70% de forma a induzir a sua morte, e a remover substâncias adiposas presentes na superfície corporal. No total, foram reunidos dois lotes de formigas para diferentes fins. O primeiro em meados de Novembro serviu para análise dos genes do rRNA 16S, e o segundo em meados de Abril para amplificação dos genes alvo através de primers “consenso” degenerados. Ademais, o procedimento de cada amostra foi efetuado com equipamento estéril.

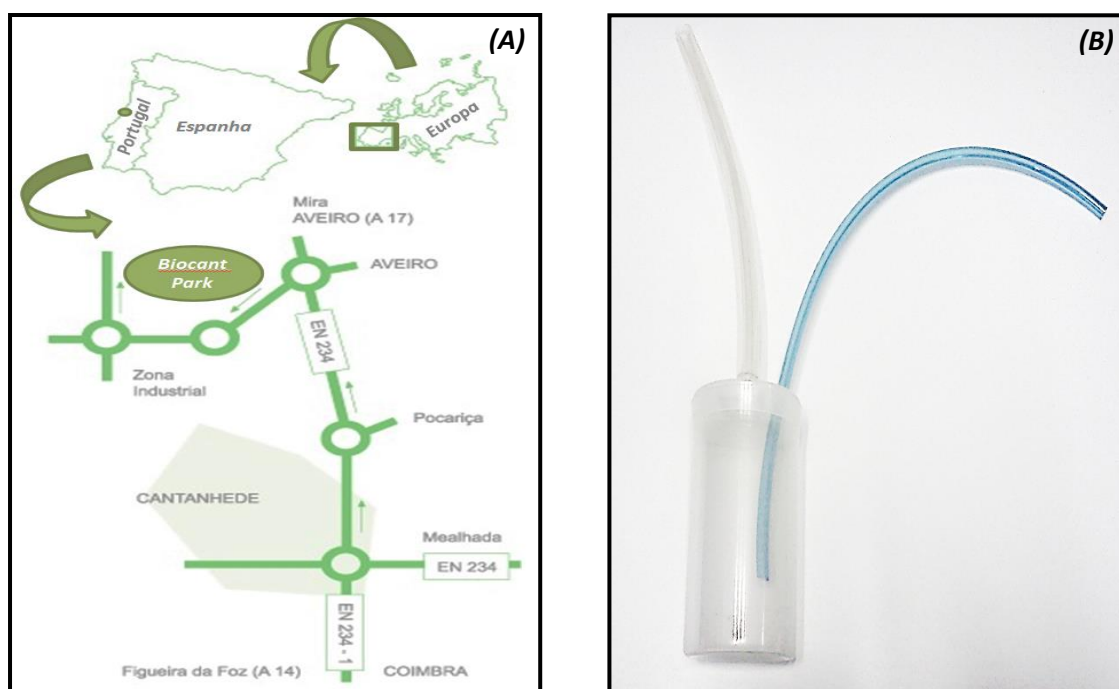


Figura 17 – Localização geográfica do local onde se efetuou a recolha das formigas (A) adaptado [105] e imagem do “Pooter” – apanhador de formigas (B).

2. Dissecção do interior do gáster das formigas

Inicialmente, a superfície externa das formigas foi esterilizada com lixívia a 5%, num tubo Falcon de 50 mL, a fim de eliminar todos os microrganismos presentes na superfície externa das formigas. Posteriormente, as formigas foram lavadas algumas vezes em água destilada para remover a lixívia.

Para ambos os lotes de formigas, o interior do gáster foi dissecado com o auxílio de uma pinça e um bisturi. Este processo foi efetuado no Microscópio Estereoscópico Leica MZ75 (Leica Microsystems, Wetzlar, Switzerland). No primeiro lote, o interior do gáster de aproximadamente 60 formigas foi removido para um tubo eppendorf que continha 200 µL de tampão ASL (QIAGEN, QIAamp DNA Stool), e foi processado no mesmo dia. Por sua vez, no segundo lote utilizaram-se cerca de 600 formigas e o interior do gáster foi armazenado em álcool a 70% à temperatura de 4 °C para posterior utilização, uma vez que este procedimento é moroso e de forma a garantir que as bactérias e o seu DNA não se deteriorassem.

3. Extração do DNA da comunidade microbiana

Para cada lote de formigas referido anteriormente, a extração de DNA da comunidade microbiana foi realizada em separado e com o kit, QIAamp DNA Stool Mini Kit (QIAGEN), de acordo com o procedimento do manual e com ligeiras alterações [104]. O procedimento deste kit é rápido, pouco complexo e permite o isolamento e a purificação da totalidade de DNA presente na amostra.

De uma forma geral, o kit engloba as seguintes etapas: lise das amostras em tampão ASL, adsorção de impurezas por pastilhas InhibitEX e purificação do DNA em colunas QIAamp spin. Além do mais, o seguinte procedimento de extração de DNA foi ajustado à primeira amostra que continha o interior do gáster de aproximadamente 60 formigas.

Em primeiro lugar procedeu-se à lise das células por homogeneização mecânica com esferas de vidro. As esferas de vidro (Ø 0,5-0,75 mm; RETSCH, Haan, Germany) foram hidratadas com água Milli-Q (Merck Millipore, Millipore Synthesis A10) e transferidas (200 µL) para o eppendorf que continha o interior do gáster das formigas, juntamente com 500 µL de tampão ASL (QIAGEN, QIAamp DNA Stool Mini Kit). A homogeneização

efetuou-se no Tissue-lyser (QIAGEN, Hilden, Germany) com 2 ciclos de 5 minutos, a 30 Hz, com 2 minutos em gelo entre os ciclos. A amostra foi colocada num bloco térmico durante 5 minutos a 95 °C e de seguida centrifugada à velocidade máxima durante 1 minuto (14800 rpm, Microcentrifuge Sigma-Aldrich 1-14). O sobrenadante foi recolhido e colocado num novo tubo.

Na segunda etapa, foi adicionada ½ pastilha InhibitEX à amostra e levada de imediato ao vórtex durante aproximadamente 1 minuto, até a pastilha se desfazer. De seguida, a amostra foi incubada durante 1 minuto à temperatura ambiente, de forma a permitir a adsorção dos inibidores de PCR à matriz InhibitEX. A amostra foi de novo centrifugada à velocidade máxima (14800 rpm, Microcentrifuge Sigma-Aldrich 1-14) durante 3 minutos, o sobrenadante (~300 µL) recolhido para um novo tubo que continha 300 µL de tampão AL e 22,5 µL de Proteinase K a 20 mg/mL, e os “pellets” descartados. A homogeneização foi realizada no vórtex durante 15 segundos e a amostra foi incubada a 70 °C durante 10 minutos. Nesta fase, a degradação do DNA é inativada uma vez que se procede à inativação de DNases e RNases. Posteriormente, o DNA foi precipitado com 300 µL de etanol a 100%.

Por último, a amostra foi transferida em frações de 500 µL para uma coluna QIAamp Mini Spin (QIAGEN, DNeasy Blood & Tissue Kit) e centrifugada à velocidade máxima durante 1 minuto (14800 rpm, Microcentrifuge Sigma-Aldrich 1-14). Este passo foi novamente repetido até se adicionar a totalidade da amostra à coluna. A coluna foi colocada num novo tubo coletor e adicionaram-se 500 µL de tampão AW1 (QIAGEN, QIAamp DNA Stool Mini Kit). Após centrifugação à velocidade máxima por 1 minuto (14800 rpm, Microcentrifuge Sigma-Aldrich 1-14), o tubo coletor com o filtrado foi descartado. A amostra foi de novo lavada com 500 µL de tampão AW2 (QIAGEN, QIAamp DNA Stool Mini Kit), centrifugada à velocidade máxima durante 2 minutos e o tubo coletor descartado. A coluna foi colocada num novo tubo e adicionaram-se 200 µL de tampão AE (QIAGEN, QIAamp DNA Stool Mini Kit). A amostra foi incubada à temperatura ambiente durante 5 minutos e centrifugada a 10000 rpm durante 1 minuto (Microcentrifuge Sigma-Aldrich 1-14).

Após purificação do DNA, as duas amostras foram concentradas no Savant SpeedVac® DNA 120 Concentrator (Thermo Fisher Scientific, Waltham, MA, USA) para reduzir o volume e armazenadas a -20 °C para posterior utilização. A quantidade de DNA

da amostra foi verificada através do NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

4. Estudo da comunidade bacteriana – Pirosequenciação e análise dos genes do rRNA 16S

4.1 Amplificação por PCR da região V3V4 dos genes do rRNA 16S

Após extração do DNA, para a primeira amostra correspondente ao DNA do interior do gáster de aproximadamente 60 formigas, a região V3V4 dos genes do rRNA 16S ribossomal foi amplificada por PCR recorrendo à utilização dos primers descritos na Tabela 1.

Tabela 1 – Sequências dos primers utilizados na amplificação por PCR da região V3V4 do genes do rRNA 16S.

Direção	Sequência do primer (5'→3')
V3 Forward	ACT CCT ACG GGA GGC AG
V4 Reverse	TAC NVR RGT HTC TAA TYC

Para além das sequências específicas para a região V3V4 (Tabela 1), os primers “forward” e “reverse” continham nas extremidades 5’ adaptadores do tipo A e do tipo B, respetivamente, adaptadores estes necessários para o processo de pirosequenciação, e uma sequência diferenciadora de 8 nucleótidos, o marcador (SD), entre o adaptador A e o primer V3, de forma a permitir a identificação das respetivas sequências geradas na pirosequenciação (Figura 18).

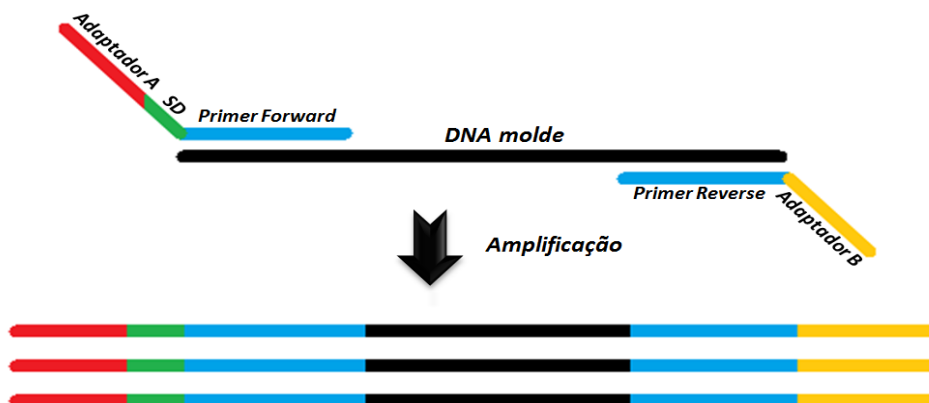


Figura 18 – Representação da composição dos primers “forward” e “reverse” e dos fragmentos amplificados dos genes do rRNA 16S.

Para a amplificação foi preparada uma mistura contendo: 5% de DMSO (v/v) (Roche Diagnostics GmbH, Mannheim, Germany); 1x de tampão Advantage® 2 SA PCR (Clontech Laboratories, Inc., Mountain View, CA, USA); 1x Advantage® 2 Polymerase Mix; 0,2 mM de mistura de dNTPs (Bioron, Ludwigshafen am Rhein, Germany); 0,2 µM de cada primer (Sigma-Aldrich, Missouri, US); e 2 µL de DNA molde. O restante volume foi completado com água Milli-Q (Merck Millipore, Millipore Synthesis A10) até perfazer 25 µL. Como controlo positivo foi utilizado DNA genómico de *Rubrobacter sp.* (25 ng/µL) e como controlo negativo a mesma mistura sem DNA.

A amplificação por PCR foi efetuada no termociclador MyCycler (Bio-Rad Laboratories, Hercules, CA, USA) e realizada sob as seguintes condições: desnaturação inicial a 94 °C durante 4 minutos; 35 ciclos de desnaturação a 94 °C durante 30 segundos; emparelhamento a 44 °C durante 45 segundos; alongamento a 68 °C durante 60 segundos; e um passo de extensão final a 68 °C durante 10 minutos.

O tamanho e qualidade dos produtos de PCR foram verificados por eletroforese em gel de agarose a 1% (w/v) com 1,2 µL de brometo de etídio 10 mg/mL (Electran®, Leuven, Belgium). A cada amostra de 5 µL foi adicionado 1 µL de GelPilot Loading Dye 5x (QIAGEN, Mainz, Germany), cujas foram corridas durante 20 minutos a 90 Volts. Em cada gel foi incluído 1,5 µL de um padrão NZYDNA Ladder I (NZYTech, Lisboa, Portugal). As imagens dos géis de agarose foram recolhidas através do Universal Hood II Gel Imager (Bio-Rad Laboratories, Milan, Italy).

4.2 Pirosequenciação

O amplicão da região V3V4 dos genes do rRNA 16S gerado foi purificado com “beads” AMPure XP (Agencourt, Beckman Coulter, USA) conforme o manual de instruções [105], para a remoção de excesso de primers e o grau de pureza avaliado em gel de agarose a 1% (w/v).

O DNA presente na amostra foi quantificado por fluorimetria, através do kit QuantiT™ PicoGreen® dsDNA (Invitrogen, Carlsbad, CA, USA) de acordo com o procedimento do manual [106]. De seguida, o número de moléculas por microlitro presente na amostra foi determinado pela fórmula apresentada abaixo:

$$\text{N}^{\circ} \text{ moléculas}/\mu\text{L} = \frac{\text{concentração amostra} \left(\frac{\text{ng}}{\mu\text{L}} \right) \times \text{constante de Avogadro}}{\text{peso molecular médio das bases} \times \text{tamanho amplicão (pb)}}$$

Para determinar a percentagem de eficiência de amplificação, a amostra foi titulada por intermédio do kit KAPA Library Quantification for Roche 454 GS FLX platform (Kapa Biosystems, Boston, MA, USA) conforme o manual de instruções [107]. O PCR de emulsão foi realizado segundo o manual de instruções do fabricante (454 Life Sciences, Roche, Branford, CT, USA), cujas etapas principais foram: preparação da mistura para amplificação, captura da biblioteca de DNA, emulsificação, amplificação, recuperação das “beads”, enriquecimento das “beads” que continham o DNA amplificado e emparelhamento do primer de sequenciação [108].

Por fim, a amostra foi pirosequenciada no GS FLX Titanium (Roche, Branford, CT, USA) no Biocant (Cantanhede, Portugal), a partir do adaptador A (GS Titanium PicoTiterPlate 70x75).

4.3 Análise de dados

Finalizada a sequenciação, os dados fornecidos pelo sequenciador foram analisados pela “pipeline” de classificação automática implementada na Unidade de Serviços Avançados do Biocant (Cantanhede, Portugal). A “pipeline” foi desenvolvida na linguagem de programação Python e associa diversos softwares, o que permite uma análise bastante rápida de sequências de rDNA.

Inicialmente, separaram-se as sequências correspondentes às diferentes amostras através da sequência diferenciadora. Em seguida, aplicaram-se critérios distintos a fim de eliminar sequências de baixa qualidade, mais propriamente: remover sequências com menos de 150 nucleótidos; excluir sequências com mais de 2 nucleótidos indefinidos; cortar o primer “reverse” no fim, caso este seja alcançado pela sequência; cortar a sequência no local onde a média de qualidade for inferior a 15 numa janela de 7 nucleótidos; remover sequências que apresentem mais de 50% de regiões de baixa complexidade - software DustMasker [109]; e eliminar quimeras - software Uchime [110].

Posteriormente, no software Usearch [111] procedeu-se à criação de OTU, onde se agruparam as sequências a uma distância (medida de diferença entre sequências) de 3%

que corresponde ao limite a nível de espécie. Além disso, foi calculado o índice de diversidade (Chao1) e a riqueza da população (curvas de rarefação) por intermédio do software Mothur [112]. As OTUs foram comparadas com as sequências presentes na base de dados Ribosomal Database Project II (RDP) [96], através da ferramenta Blastn (Basic Local Alignment Search Tool) para identificar a taxonomia das mesmas.

Os “hits” obtidos anteriormente foram submetidos a um novo controlo de qualidade, rejeitando-se aqueles com uma similaridade inferior a 40% e com um “E-value” superior a $1e^{-50}$. Ao mesmo tempo, foram aplicados testes de “bootstrap” com valores superiores a 70%, a fim de confirmar o grau de confiança relativamente à identificação taxonómica das OTUs. Este ensaio foi executado por intermédio do software seqBoot do Phylip [113], onde são concebidas 100 réplicas com 5% de variações em relação às OTUs em análise.

Todo este procedimento de identificação taxonómica é extremamente complexo, dado que é indispensável corrigir a pontuação “E-value” e examinar todos os níveis taxonómicos, incluindo simultaneamente o teste “bootstrap”.

5. Determinação de genes alvo codificantes de enzimas

5.1 Desenho dos primers “consenso” degenerados

Os primers “consenso” degenerados foram desenhados de forma a emparelharem em regiões conservadas de genes que codificam as enzimas cutinases, feruloil esterases e lacases. É também importante referir que, estas regiões deverão localizar-se junto às extremidades dos genes, a fim de os primers amplificarem o maior número de nucleótidos possível.

Inicialmente, identificaram-se os aminoácidos extremamente conservados nas enzimas alvo, uma vez que a ausência ou alteração destes aminoácidos pode significar a perda da atividade por parte das enzimas em questão. Estas regiões “consenso” foram identificadas recorrendo à base de dados Pfam (Protein Families Database) [114], na qual é possível reconhecer aminoácidos de regiões conservadas de numerosas classes específicas de enzimas, por intermédio de alinhamentos múltiplos de sequências.

Para aumentar a especificidade de amplificação dos primers, encontraram-se sequências de genes das enzimas alvo correspondentes aos géneros e famílias identificadas pela pirosequenciação dos genes do rRNA 16S, ou seja, com o intuito de os primers

amplificarem eficientemente as enzimas bacterianas presentes na amostra do gáster das formigas. Desta forma, utilizou-se a base de dados UniProt (Universal Protein Resource) [115] como motor de busca para estas sequências.

Por sua vez, como os primers são contruídos em nucleótidos, na base de dados ENA (European Nucleotide Archive) da EMBL-EBI [116] obtiveram-se as respetivas sequências em nucleótidos que se encontravam em aminoácidos na base de dados UniProt.

Seguidamente, efetuaram-se alinhamentos múltiplos das sequências obtidas para cada uma das enzimas alvo em separado. O alinhamento múltiplo permite a visualização e confirmação das regiões conservadas verificadas anteriormente na base de dados Pfam. As sequências de nucleótidos foram introduzidas no software MEGA 5 (Molecular Evolutionary Genetics Analysis) [117] onde se realizaram alinhamentos múltiplos de sequências em aminoácidos por intermédio do MUSCLE [101]. Além disso, este software permite a conversão de sequências de nucleótidos para aminoácidos e vice-versa.

Posteriormente à identificação de cada região conservada das enzimas alvo, foram excluídas as sequências que não continham as respetivas regiões conservadas, assim como, aquelas que apresentavam menor similaridade nessa mesma zona “consenso”. Desta forma, para que cada primer pudesse ser contruído, verificou-se se cada região “consenso” apresentava aproximadamente 20 nucleótidos consecutivos, o equivalente a 7 aminoácidos.

Para permitir cobrir aminoácidos extremamente conservados nas enzimas alvo de diferentes grupos microbianos, na construção dos primers “consenso” foram incluídos no máximo 7 nucleótidos degenerados, isto é, para uma localização específica de um nucleótido pode corresponder dois ou mais nucleótidos. As bases degeneradas presentes nos primers foram elaboradas de acordo com designação atribuída pela IUB (International Union of Biochemistry).

É de notar que, cada primer “forward” corresponde à região conservada localizada junto à extremidade 5’ do gene e tem como sequência essa mesma região, enquanto que o primer “reverse” corresponde à região “consenso” localizada junto à extremidade 3’ do gene e a sua sequência é definida pelas suas bases complementares, convertidas em sentido inverso.

Após a construção, averiguaram-se as propriedades de cada primer tendo em conta os seguintes parâmetros: a diferença de temperatura de melting de cada par de primers não ser superior 10 °C, o conteúdo em GC ser de preferência entre 40% e 60%, e self annealing e loops reduzidos. Todos estes parâmetros foram avaliados no software Oligo Analyser 1.5.

Por último, para verificar a especificidade dos primers para microrganismos e descartar amplificações inespecíficas a partir da formiga utilizou-se a ferramenta Blastn do NCBI (National Center for Biotechnology Information) [118] e da base de dados ant genomics (Fourmidable) [119].

Os primer construídos para cada uma das enzimas alvo estão descritos na Tabela 2, e para além disso, todo o procedimento encontra-se esquematizado na Figura 19 para uma melhor compreensão.

Tabela 2 – Lista dos primers “consenso” degenerados construídos para cutinases, feruloil esterases e lacases, tendo em conta a designação atribuída para as bases degeneradas pela IUB.

Cutinases					
Nome_Direção	Sequência do primer (5'→3')	Comprimento (pb)	Bases degeneradas	TM média (°C)	GC médio (%)
Cut_F1	TBC CBG GNA CVT GGG ARA C	19	5	59,8	63,2
Cut_F2	TAT TWT BTR ACV GAT ACR CC	20	5	47,8	36,7
Cut_R1	GSV GTS GMG CAG TAS AGR TC	20	6	59,3	63,3
Cut_R2	GAT YTG VAG RTT SAG CCA RTC	21	5	54,0	48,4
Cut_R3	KAT CGW MTG CTK SAA WAT YT	20	7	48,6	35,0
Feruloil esterases					
Nome_Direção	Sequência do primer (5'→3')	Comprimento (pb)	Bases degeneradas	TM média (°C)	GC médio (%)
Fae_F1	GGN TGC TCN RVN GGN GGN C	19	7	62,9	71,9
Fae_F2	GGN TGY TCN GAY GGN GGN MG	20	7	61,9	67,5
Fae_R1	GGR TAN NBG CAN AGN GGN CG	20	7	59,7	63,3
Fae_R2	GCN HBG CCN TKY TCN ACC CA	20	7	61,7	62,5
Lacases					
Nome_Direção	Sequência do primer (5'→3')	Comprimento (pb)	Bases degeneradas	TM média (°C)	GC médio (%)
Lac_F1	ACN WCN RTN CAY TGG CAY GG	20	7	57,9	55,0
Lac_R1	TGN TCN AGD ATR TGR CAR TG	20	6	51,7	44,2
Lac_R2	WCR TGR CAR TGR AAC ATC CA	20	5	54,4	45,0
Lac_R3	AGR TGR CAR TGV HAN GCC CA	20	6	59,4	55,0

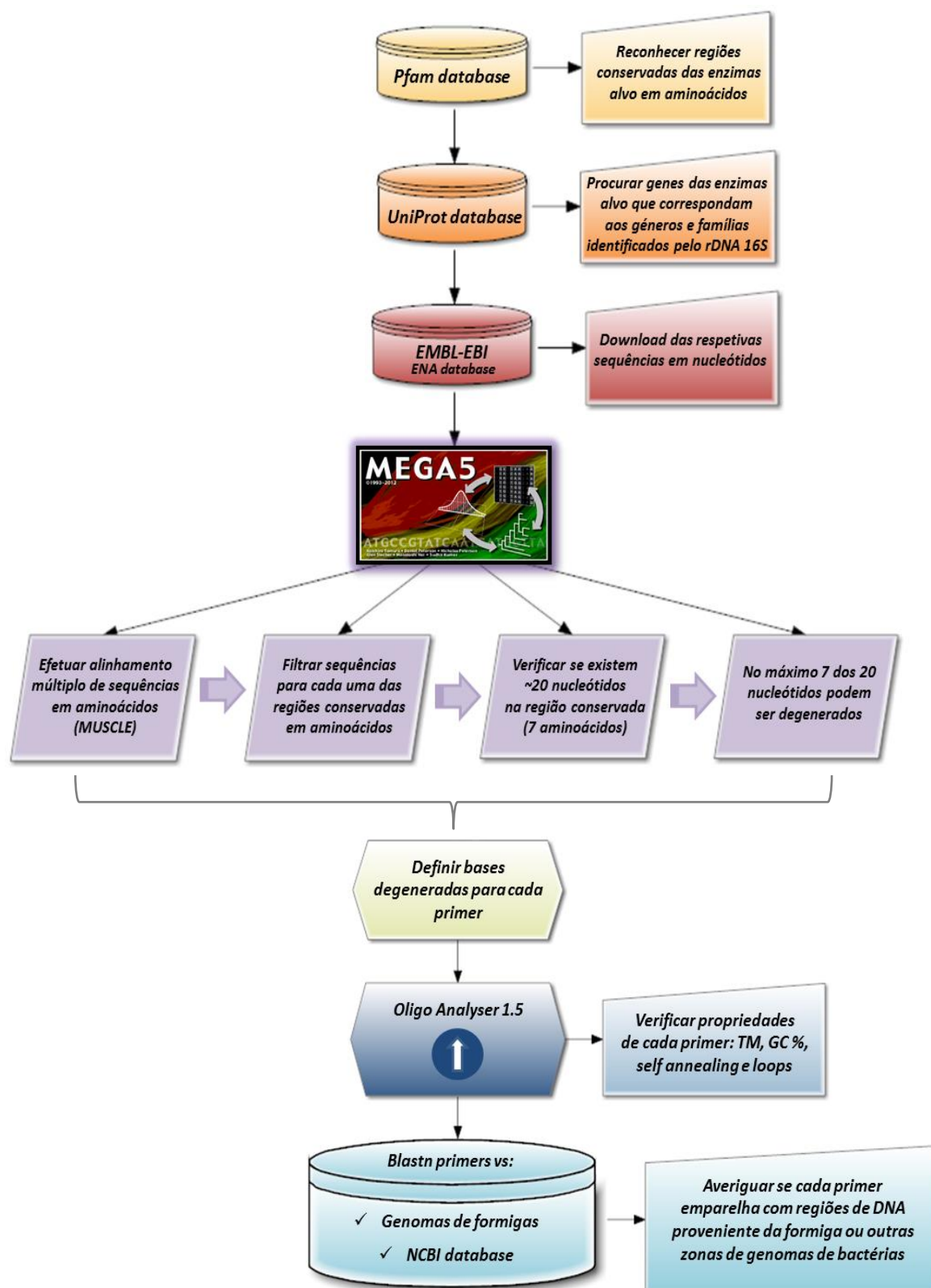


Figura 19 – Procedimento realizado para a construção de primers “consenso” degenerados.

5.2 Amplificação por PCR através de primers “consenso” degenerados

A seguir à construção dos primers “consenso” degenerados e extração do DNA da segunda amostra, correspondente ao DNA do interior do gáster de aproximadamente 600 formigas, procedeu-se à amplificação por PCR das zonas localizadas entre as regiões “consenso” de genes que codificam as diferentes classes de enzimas: cutinases, feruloil esterases e lacases.

Para as respetivas classes de enzimas foram realizadas amplificações que envolveram um ou mais do que um par de primers na mesma reação, ou seja, utilizaram-se diferentes pares de primers em conjunto (“pool” de primers) e/ou em separado, conforme o tamanho dos fragmentos de DNA esperados e obtidos ao longo da otimização das reações. Desta forma, a metodologia utilizada para as feruloil esterases e lacases abrangeu um “pool” de primers para cada uma das mesmas, enquanto que na amplificação das cutinases, as diferentes reações previstas foram desdobradas e, utilizado em separado apenas um par de primers por cada amplificação.

Na amplificação de cutinases foram preparadas três misturas diferentes, na qual cada uma destas continha: 5% de DMSO (v/v) (Roche Diagnostics GmbH, Mannheim, Germany); 1x de tampão Advantage® 2 SA PCR (Clontech Laboratories, Inc., Mountain View, CA, USA); 1x Advantage® 2 Polymerase Mix; 0,2 mM de mistura de dNTPs (Bioron, Ludwigshafen am Rhein, Germany); 0,2 µM de cada primer “forward” e “reverse” (Integrated DNA Technologies, Munich, Germany); e 3 µL de DNA molde. O restante volume foi completado com água Milli-Q (Merck Millipore, Millipore Synthesis A10) até perfazer 20 µL. Os pares de primers utilizados nas respetivas amplificações em separado foram os seguintes: 1ª) Cut_F1 com Cut_R1; 2ª) Cut_F1 com Cut_R2; 3ª) Cut_F2 com Cut_R3 (Tabela 2). Como controlo negativo foi utilizado a mesma mistura sem DNA.

As diferentes amplificações por PCR de cutinases foram efetuadas no termociclador MyCycler (Bio-Rad Laboratories, Hercules, CA, USA) e realizada conforme as seguintes condições:

- ✓ 1ª) desnaturação inicial a 94 °C durante 4 minutos; 35 ciclos de desnaturação a 94 °C durante 30 segundos; emparelhamento a 54 °C durante 45 segundos;

alongamento a 68 °C durante 90 segundos; e um passo de extensão final a 68 °C durante 10 minutos;

- ✓ 2ª) desnaturação inicial a 94 °C durante 4 minutos; 35 ciclos de desnaturação a 94 °C durante 30 segundos; emparelhamento a 48 °C durante 45 segundos; alongamento a 68 °C durante 150 segundos; e um passo de extensão final a 68 °C durante 10 minutos;
- ✓ 3ª) desnaturação inicial a 94 °C durante 4 minutos; 35 ciclos de desnaturação a 94 °C durante 30 segundos; emparelhamento a 52 °C durante 45 segundos; alongamento a 68 °C durante 150 segundos; e um passo de extensão final a 68 °C durante 10 minutos.

Relativamente à amplificação de feruloil esterases foi preparada uma mistura contendo: 5% de DMSO (v/v) (Roche Diagnostics GmbH, Mannheim, Germany); 1x de tampão Advantage® 2 SA PCR (Clontech Laboratories, Inc., Mountain View, CA, USA); 1x Advantage® 2 Polymerase Mix; 0,2 mM de mistura de dNTPs (Bioron, Ludwigshafen am Rhein, Germany); 0,2 µM de cada primer (Integrated DNA Technologies, Munich, Germany); e 3 µL de DNA molde. O restante volume foi completado com água Milli-Q (Merck Millipore, Millipore Synthesis A10) até perfazer 20 µL. O conjunto de primers utilizados foram Fae_F1, Fae_F2, Fae_R1 e Fae_R2 (Tabela 2). Como controlo negativo foi utilizado a mesma mistura sem DNA.

A amplificação por PCR de feruloil esterases foi efetuada no termociclador MyCycler (Bio-Rad Laboratories, Hercules, CA, USA) e realizada sob as seguintes condições: desnaturação inicial a 94 °C durante 4 minutos; 35 ciclos de desnaturação a 94 °C durante 30 segundos; emparelhamento a 62 °C durante 45 segundos; alongamento a 68 °C durante 120 segundos; e um passo de extensão final a 68 °C durante 10 minutos.

Quanto à amplificação por PCR de lacases foi preparada uma mistura que continha: 5% de DMSO (v/v) (Roche Diagnostics GmbH, Mannheim, Germany); 1x de tampão Advantage® 2 SA PCR (Clontech Laboratories, Inc., Mountain View, CA, USA); 1x Advantage® 2 Polymerase Mix; 0,2 mM de mistura de dNTPs (Bioron, Ludwigshafen am Rhein, Germany); 0,4 µM do primer “forward” (Integrated DNA Technologies, Munich, Germany); 0,2 de cada primer “reverse” (Integrated DNA Technologies, Munich, Germany); e 3 µL de DNA molde. O restante volume foi completado com água Milli-Q (Merck Millipore, Millipore Synthesis A10) até perfazer 20 µL. A “pool” de primers

continha: Lac_F1, Lac_R1, Lac_R2 e Lac_R3 (Tabela 2). Como controlo negativo foi utilizado a mesma mistura sem DNA.

A amplificação por PCR de lacases foi efetuada no termociclador MyCycler (Bio-Rad Laboratories, Hercules, CA, USA) e realizada conforme as seguintes condições: desnaturação inicial a 94 °C durante 4 minutos; 35 ciclos de desnaturação a 94 °C durante 30 segundos; emparelhamento a 58 °C durante 45 segundos; alongamento a 68 °C durante 120 segundos; e um passo de extensão final a 68 °C durante 10 minutos.

É de notar que, para se obter DNA suficiente para que a biblioteca de amplicões fosse sequenciada (500 ng) foi necessário realizar no total onze amplificações: duas para 1ª de cutinases (Cut_F1 com Cut_R1); duas para 2ª de cutinases (Cut_F1 com Cut_R2); uma para 3ª de cutinases (Cut_F2 com Cut_R3); três para feruloil esterases (“pool” de primers); e três para lacases (“pool” de primers).

O tamanho e qualidade dos produtos de PCR foram verificados por eletroforese em gel de agarose a 1% (w/v) com 1,2 µL de brometo de etídio 10 mg/mL (Electran®, Leuven, Belgium). A cada amostra de 5 µL foi adicionado 1 µL de GelPilot Loading Dye 5x (QIAGEN, Mainz, Germany). A eletroforese decorreu durante 20 minutos a 90 Volts. Em cada gel foi incluído 1,5 µL de um padrão NZYDNA Ladder I (NZYTech, Lisboa, Portugal). As imagens dos géis de agarose foram obtidas através do Universal Hood II Gel Imager (Bio-Rad Laboratories, Milan, Italy).

5.3 Pirosequenciação

Subsequentemente à amplificação por PCR pelos primers “consenso” degenerados, todos os produtos de PCR gerados foram purificados com “beads” AMPure XP (Agencourt, Beckman Coulter, USA) conforme o manual de instruções [105], para a remoção de excesso de primers e o grau de pureza avaliado em gel de agarose a 1% (w/v).

A amplificação por PCR pelos primers “consenso” degenerados gerou amplicões de diferentes tamanhos correspondentes às classes distintas de enzimas alvo. Tendo em conta que, os amplicões deverão ter entre 200 a 600 pb, requeridos na emulsão de PCR da pirosequenciação, os que apresentaram um tamanho superior a 800 pb foram fragmentados no Ultrasonicador S220 (Covaris, Woburn, MA, USA) sob as seguintes condições: “peak incident power” de 105 W; “duty factor” de 5%; 200 “cycles per burst”; e um tempo de

120 segundos. Todas as amplificações foram sujeitas a ultrasonificação excluindo a 1ª de cutinases, dado que foi a única que gerou fragmentos de DNA inferiores a 800 pb.

O DNA presente em cada amostra foi quantificado por fluorimetria, através do kit Quant-iT™ PicoGreen® dsDNA (Invitrogen, Carlsbad, CA, USA) conforme o manual de instruções [106]. De seguida, o número de moléculas por microlitro presente em cada amostra foi determinado pela fórmula referida anteriormente (Secção 4.2). Desta forma, realizou-se uma mistura equimolar dos diferentes fragmentos correspondentes às enzimas alvo, para que cada amostra estivesse representada com número semelhante de moléculas na mistura final (biblioteca de enzimas).

Seguidamente, efetuaram-se as seguintes etapas de acordo com as instruções do manual do fabricante Rapid Library Preparation Method Manual (454 Life Sciences, Roche, Branford, CT, USA): reparação e adenilação dos terminais dos fragmentos quebrados por ultrasonificação; preparação das “beads” AMPure; ligação de adaptadores; remoção de fragmentos pequenos; e quantificação da biblioteca [120]. Para uma melhor compreensão, na Figura 20 está representado todo este procedimento que antecede a entrada da biblioteca de enzimas no pirosequenciador. É também importante referir que, a partir dos adaptadores o processo de sequenciação é iniciado de forma aleatória, ou seja, desta forma pode ocorrer sequenciação de ambos os lados dos amplicões, permitindo cobrir fragmentos de tamanho superiores aos recomendados na pirosequenciação.

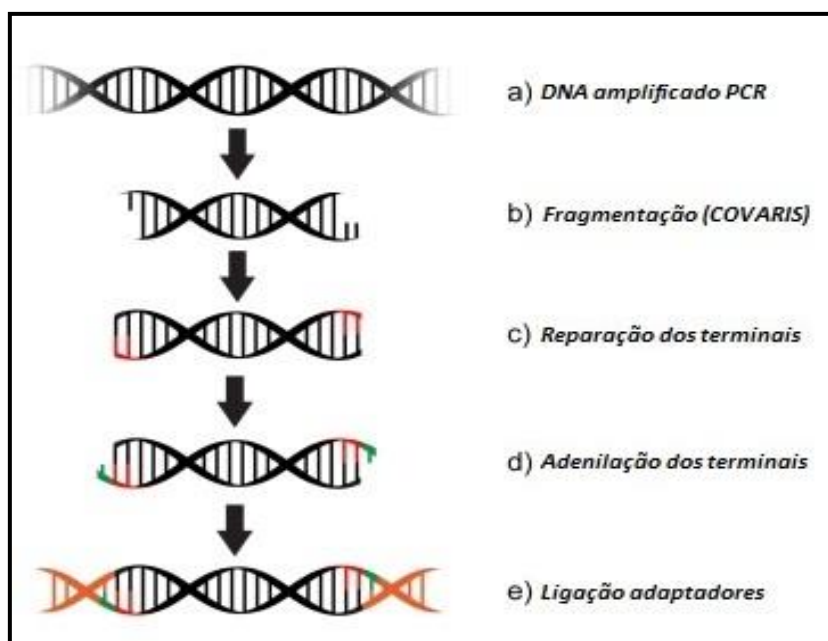


Figura 20 – Sumário do procedimento realizado para a biblioteca de enzimas alvo.

Para determinar a percentagem de eficiência de amplificação, a amostra (“pool” de amplicões) foi titulada por intermédio do kit KAPA Library Quantification for Roche 454 GS FLX platform (Kapa Biosystems, Boston, MA, USA) de acordo com o procedimento do manual [107]. O PCR de emulsão foi realizado segundo o manual de instruções do fabricante (454 Life Sciences, Roche, Branford, CT, USA), cujas etapas principais foram: preparação da mistura para amplificação, captura da biblioteca de DNA, emulsificação, amplificação, recuperação das “beads”, enriquecimento das “beads” que continham o DNA amplificado e emparelhamento do primer de sequenciação [121].

Por fim, a amostra foi pirosequenciada no GS FLX Titanium (Roche, Branford, CT, USA) no Biocant (Cantanhede, Portugal), a partir dos adaptadores presentes em ambas as extremidades nos amplicões (GS Titanium PicoTiterPlate 70x75).

5.4 Análise de dados

Após sequenciação na plataforma 454 da Roche, as sequências obtidas foram processadas para eliminação dos adaptadores e submetidas à “pipeline” de análise MG-RAST. Esta “pipeline” permite a anotação funcional e taxonómica de sequências de amostras metagenómicas por comparação com bases de dados como: Genebank, eggNOG, KEGG, PATRIC, RefSeq, SwissProt, SEED e TrEMBL.

III - RESULTADOS E DISCUSSÃO

1. Composição do microbioma intestinal da formiga

A composição da comunidade microbiana do intestino de *C. scutellaris* foi estudada através da construção de uma biblioteca de amplicões da região hipervariável V3V4 dos genes do rRNA 16S bacteriano. A biblioteca de amplicões foi pirosequenciada recorrendo à tecnologia 454 e a taxonomia identificada por comparação com dados presentes em bases de dados curadas.

1.1 Amplificação por PCR da região V3V4 dos genes do rRNA 16S

Os produtos de PCR, resultantes da amplificação da região hipervariável V3V4 dos genes do rRNA 16S, podem ser observados em gel de agarose na Figura 21. Como esperado, obteve-se um amplicão com um tamanho entre 400 e 560 pb. Além disso, observa-se uma banda inespecífica de baixa intensidade que não foi valorizada.

Deste modo, verifica-se uma boa eficiência na amplificação da região V3V4, o que reflete simultaneamente uma extração de DNA apropriada para bactérias e livre de inibidores. Por fim, a banda correspondente ao amplicão da região V3V4 foi purificada para remoção de excesso de primers e, posteriormente, sequenciada na plataforma 454 da Roche, para identificação taxonómica.

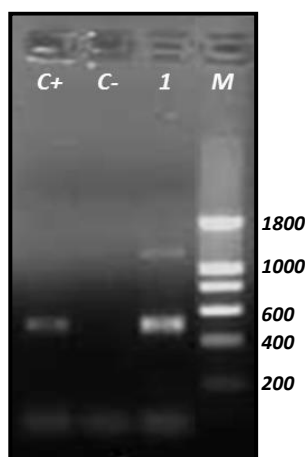


Figura 21 – Verificação do tamanho e qualidade do amplicão obtido para a região V3V4 dos genes do rRNA 16S. C+: controlo positivo (DNA genómico de *Rubrobacter sp.*); C-: controlo negativo; 1: produtos de PCR da região V3V4 do rDNA 16S; M: marcador NZYDNA Ladder I com o tamanho das bandas está representado do lado direito.

1.2 Pirosequenciação e análise da riqueza microbiana

Na Tabela 3 estão representados os resultados obtidos após o processamento de dados na pirosequenciação dos genes do rRNA 16S. No total foram geradas 12471 sequências, nas quais foram aplicadas vários parâmetros de controlo de qualidade. Destas, 93,7% passaram nos filtros de qualidade, sendo desta forma consideradas sequências de qualidade elevada.

Por sua vez, as sequências de alta qualidade (11682) foram agrupadas em 95 OTUs, onde a cada OTU pertencem sequências com uma distância inferior a 3%. Deste modo, podemos concluir que quase todas as sequências e OTUs passaram nos controlos de qualidade, uma vez que as suas percentagens correspondem a valores bastante elevados, 93,7% e 96,9%, respetivamente.

Tabela 3 – Números e percentagens de sequências e OTUs gerados no processamento de dados na pirosequenciação, assim como a percentagem de cobertura da diversidade com base nos valores de Chao1.

Amostra	# Total de sequências	# Sequências de qualidade elevada	% Sequências de qualidade elevada	Chao1 estimado	% Cobertura da diversidade
Interior do gáster (Formigas)	12471	11682	93,7	121,3	78,3
	# Total de OTUs	# OTUs de qualidade elevada	% OTUs de qualidade elevada	Chao1 obtido	
	98	95	96,9	95	

Na Figura 22 está representada a curva de rarefação, correspondente à amostra do interior do gáster das formigas, a fim de avaliar a riqueza da população bacteriana. A curva de rarefação evidencia o aumento do número de sequências de alta qualidade com o aumento do número de OTUs em que foram agrupadas. Deste modo, quando a curva atinge o “plateau” (declive igual a zero) indica que toda a diversidade bacteriana presente na amostra foi caracterizada.

Para a amostra em questão, obteve-se um valor de 78,4% na cobertura da diversidade, calculada pela razão entre os valores de Chao1 obtido e Chao1 esperado (Tabela 3). Nesta medida, podemos depreender que a grande maioria da diversidade bacteriana presente na amostra foi determinada, por isso, mesmo que o número de sequências fosse superior, o número de OTUs gerados seria praticamente o mesmo.

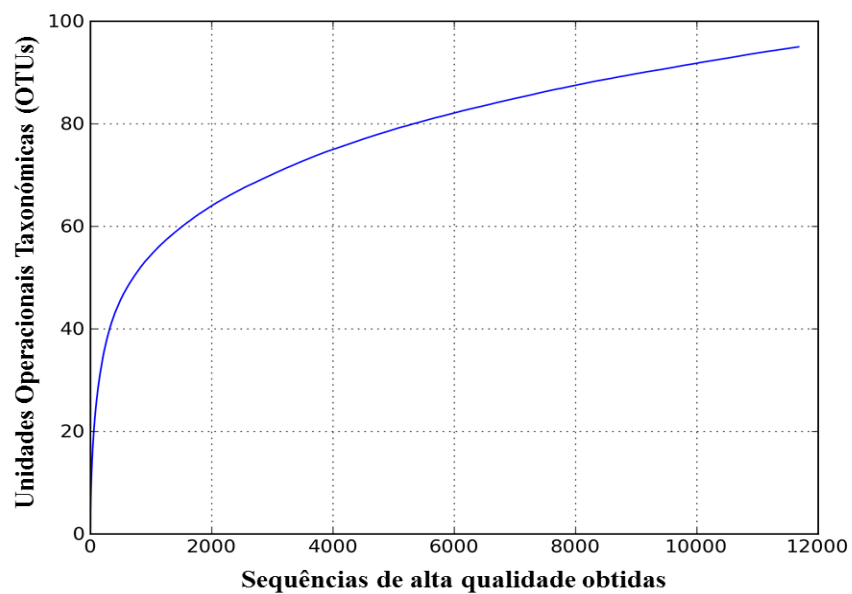


Figura 22 – Curva de rarefação com as respectivas OTUs e sequências de alta qualidade obtidas.

1.3 Identificação taxonómica bacteriana

A amostra correspondente ao interior do gáster das formigas foi caracterizada por intermédio da região V3V4 do gene do RNA ribossomal 16S, o que permitiu a identificação taxonómica no que respeita ao domínio *Bacteria*. Esta identificação teve, como base, o número de sequências de alta qualidade amplificadas para esta região (Tabela 5, Secção 1.4), cujas percentagens dos taxa vão ser descritas ao longo desta secção.

No gráfico da Figura 23 estão representados em forma circular dois taxa bacterianos, mais propriamente, o círculo externo reflete a ordem associada a cada um dos filios – o círculo interno. Verifica-se portanto, uma elevada diversidade e uma predominância bacteriana no filo *Proteobacteria* (68,0%), seguindo-se *Bacteroidetes* (17,2%), *Firmicutes* (7,0%), *Acidobacteria* (4,6%) e por último *Actinobacteria* (3,2%).

Relativamente à ordem, no filo *Proteobacteria* destacam-se *Burkholderiales* (26,9%), posteriormente *Caulobacteriales* (13,2%), *Sphingomonadales* (7,9%), *Xanthomonadales* (6,3%), *Pseudomonadales* (5,0%), *Enterobacteriales* (4,2%), *Rhizobiales* (4,1%) e *Rhodospirillales* (0,4%). Por sua vez, apenas 1,1 de 17,2% das sequências pertencentes ao filo *Bacteroidetes* foram associadas à respetiva ordem, das quais 0,6% correspondem a *Sphingobacteriales* e 0,5% a *Bacteroidales*. Quanto aos restantes filios: *Firmicutes* é composto por *Bacillales* (4,4%) e *Lactobacillales* (2,4%); *Acidobacteria* apenas por *Acidobacteriales* (3,9%); *Actinobacteria* maioritariamente por *Actinomycetales* (2,7%) e em menor número por *Solirubrobacteriales* (0,5%). É de notar que, 17,0% do total das sequências não foram associadas à respetiva ordem, entre as quais 16,1% pertencem ao filo *Bacteroidetes* – espaço em branco no círculo externo do gráfico da Figura 23.

Os resultados obtidos estão em concordância com estudos publicados previamente, nomeadamente, o principal filo presente nas várias regiões intestinais de formigas é *Proteobacteria*, de entre os quais se salientam as ordens *Rhizobiales*, *Burkholderiales* e *Xanthomonadales* [82; 122]. Além disso, microrganismos simbioses pertencentes ao filo *Bacteroidetes* encontram-se no trato gastrointestinal de animais e são bastante comuns em vários insetos, entre os quais formigas “carpinteiras” e térmitas [123; 124]. Acredita-se que a flora intestinal, especialmente *Bacteroidetes*, tem como função fornecer nutrientes aos seus hospedeiros, através da degradação de matéria orgânica de elevado peso molecular. Simultaneamente, complementam os genomas eucariotas com enzimas digestivas, cujo

alvo são polímeros dietéticos resistentes a enzimas do hospedeiro, como compostos presentes nas paredes celulares de plantas – celulose, pectinas e xilanas [124].

Outro exemplo, são os filos *Actinobacteria* e *Firmicutes* que foram já detetados em regiões intestinais de formigas [122; 125], assim como, para além destes dois, o filo *Acidobacteria* em térmitas que se alimentam de madeira [126; 127].

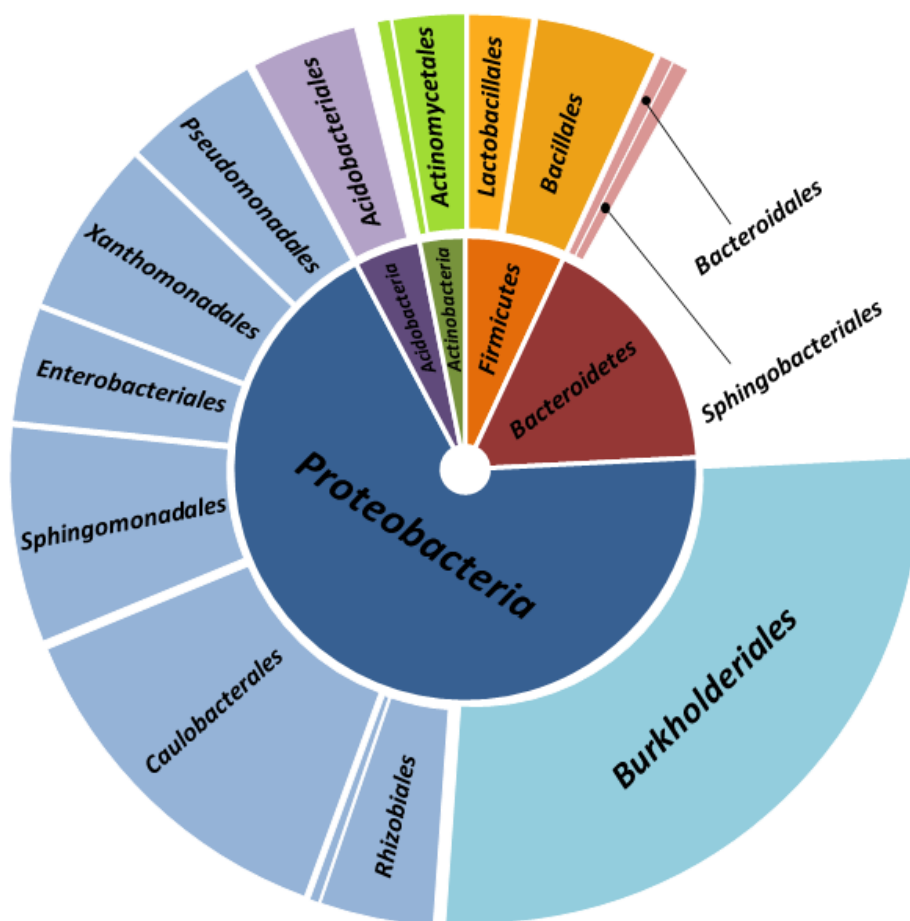


Figura 23 – Análise da distribuição taxonómica da comunidade bacteriana presente no interior do gáster das formigas com base nas sequências dos genes do rRNA 16S. O círculo interno representa a abundância a nível do filo, enquanto que o círculo externo a nível da ordem.

Mais detalhadamente, na Figura 24 está representado um gráfico em queijo para uma melhor percepção da enorme diversidade presente no filo *Proteobacteria*, o equivalente a 68,0% do total das sequências amplificadas na amostra. Pelo círculo interno podemos observar que a amostra evidencia uma distribuição homogénea ao nível das três classes: *Betaproteobacteria* (39,6%), *Alphaproteobacteria* (37,6%) e *Gammaproteobacteria* (22,8%). No círculo externo especificam-se os respetivos géneros associados a cada uma

destas classes, com exceção para grande parte da *Alphaproteobacteria*, à qual, apenas foi obtida, a nível taxonómico, a família (*Caulobacteraceae*) – círculo intermédio. Desta forma, confirma-se que pela região V3V4, para a maioria das sequências pertencentes ao filo *Proteobacteria* foi identificado o respetivo género.

Na classe *Betaproteobacteria* predominam os géneros *Massilia* (21,5%) e *Variovorax* (13,0%), seguindo-se *Ralstonia* (3,9%) e *Pelomonas* (1,2%). Por sua vez, na classe *Alphaproteobacteria* para além da família *Caulobacteraceae* que representa um total de 21,9%, encontram-se ainda os géneros *Sphingomonas* (11,6%) , *Phyllobacterium* (3,5%) e *Methylobacterium* (0,6%). Por último, a classe *Gammaproteobacteria* é composta por *Stenotrophomonas* (9,0%), *Serratia* (5,1%), *Enhydrobacter* (5,1%), *Pseudomonas* (2,1%) e *Nevskia* (0,4%).



Figura 24 – Análise da distribuição taxonómica a nível do filo Proteobacteria, presente no interior do gáster das formigas com base nas sequências dos genes do rRNA 16S. O círculo interno representa a abundância a nível da classe, o círculo intermédio a nível da família (*Caulobacteraceae*) e o círculo externo a nível do género.

Uma curiosidade adicional é que, bactérias da ordem *Rhodospirillales*, classe *Alphaproteobacteria*, estabelecem relações simbióticas com diversos insetos que dependem de uma alimentação rica em açúcares (meladas), em locais como o estômago e os túbulos de Malpighian – característica também da espécie de formigas em estudo, *C. scutellaris* [128].







As funções exercidas pelas bactérias nos seus hospedeiros são extremamente importantes, no entanto, estão longe de ser esclarecidas. Tais funções são imprescindíveis na fixação e reciclagem de azoto, no metabolismo por fornecimento de nutrientes ou por oxidação de substratos, na defesa contra outros microrganismos prejudiciais, na manutenção da homeostasia das regiões intestinais e no equilíbrio do consórcio microbiano pela fornecimento de metabolitos a outros microrganismos benéficos ao hospedeiro.

1.4 Confirmação de genes alvo em bases de dados, presentes no genoma de bactérias identificadas na flora intestinal da formiga

Tendo em conta os resultados obtidos na identificação bacteriana pela região V3V4 dos genes do rRNA 16S, recorreu-se às bases de dados UniProt e ENA para pesquisar a presença dos genes das enzimas em estudo (cutinases, feruloil esterases e lacases), nas bactérias identificadas.

Na Tabela 4 encontram-se identificados os taxa bacterianos em que as enzimas alvo foram identificadas. Podemos observar que, nos taxa que apresentaram um maior número de sequências foram encontradas principalmente feruloil esterases e lacases. Por sua vez, as cutinases foram identificadas em poucos géneros e famílias, dado que o número de sequências presentes nas bases de dados é bastante menor comparativamente às feruloil esterases e lacases.

A pesquisa de sequências foi efetuada, em primeiro lugar, ao nível do género de forma a aumentar a especificidade por parte dos primers, e só de seguida, ao nível da família e ordem. Nesta medida, as sequências identificadas foram sobretudo a nível da família e género, na qual se obteve um número considerável de sequências para cada uma das classes de enzima alvo, sequências estas que serviram para posteriores alinhamentos múltiplos e construção de primers.

Tabela 4 - Dados gerados pela “pipeline” no processamento das sequências de rDNA, correspondentes à região V3V4 do genes do RNA ribossômico 16S. Apresentam-se os taxa ordem, família e gênero identificados, assim como o número de OTUs em que foram agrupadas as respectivas sequências. Os microrganismos com um número de sequências inferior a 20 não foram incluídos por a sua abundância ser irrelevante. O número de sequências está representado por barras de acordo com a sua abundância. A coloração de fundo dos taxa representa que foram já identificadas enzimas alvo nos seus genomas. As cores associadas aos taxa são: amarelo para cutinases ; verde tropa para FAEs ; verde claro para lacases ; azul para cutinases e lacases ; laranja para FAEs e lacases ; e cor de vinho para cutinases, FAEs e lacases .

ORDEM	FAMÍLIA	GÊNERO	OTUs	Nº SEQ.
Burkholderiales	Oxalobacteraceae	Massilia	8	1688
Caulobacterales	Caulobacteraceae		6	1517
Burkholderiales	Comamonadaceae	Variovorax	3	1019
Sphingomonadales	Sphingomonadaceae	Sphingomonas	1	909
Xanthomonadales	Xanthomonadaceae	Stenotrophomonas	2	704
Acidobacteriales	Acidobacteriaceae		1	448
Bacillales	Listeriaceae	Brochothrix	1	434
Enterobacteriales	Enterobacteriaceae	Serratia	1	404
Pseudomonadales	Moraxellaceae	Enhydrobacter	5	403
Burkholderiales	Burkholderiaceae	Ralstonia	1	307
Rhizobiales	Phyllobacteriaceae	Phyllobacterium	2	273
Pseudomonadales	Pseudomonadaceae	Pseudomonas	3	166
Lactobacillales	Leuconostocaceae	Weissella	2	156
Rhizobiales			3	117
Lactobacillales	Streptococcaceae	Lactococcus	1	99
Burkholderiales	Comamonadaceae	Pelomonas	1	92
Enterobacteriales	Enterobacteriaceae		3	80
Bacillales	Staphylococcaceae	Staphylococcus	1	77
Sphingobacteriales	Chitinophagaceae	Sediminibacterium	1	68
Actinomycetales			5	66
Actinomycetales	Nocardiaceae	Rhodococcus	1	60
Bacteroidales	Porphyromonadaceae	Porphyromonas	1	58
Solirubrobacterales	Conexibacteraceae	Conexibacter	1	56
Rhizobiales	Methylobacteriaceae	Methylobacterium	1	45
Rhizobiales	Bradyrhizobiaceae		2	44
Actinomycetales	Microbacteriaceae	Amnibacterium	1	44
Rhodospirillales	Acetobacteraceae		2	41
Actinomycetales	Microbacteriaceae	Fronthabitans	1	40
Actinomycetales	Dermaococcaceae	Dermaococcus	1	29
Xanthomonadales	Sinobacteraceae	Nevskia	1	28
Actinomycetales	Kineosporiaceae	Quadrisphaera	1	26
Lactobacillales	Leuconostocaceae	Leuconostoc	1	23
Actinomycetales	Microbacteriaceae	Subtercola	1	20

2. Pesquisa e identificação de genes de suberinases

Conforme os resultados dos géneros e famílias identificados anteriormente pela região V3V4 do rDNA 16S, foram desenhados primers “consenso” degenerados a partir de alinhamentos múltiplos de sequências presentes em bases de dados. Os genes de suberinases da comunidade microbiana do intestino de *C. scutellaris* foram amplificados por PCR através dos primers “consenso” degenerados.

Os amplicões gerados com tamanho superior a 600 pb foram fragmentados por ultrasonificação. Após sequenciação na plataforma 454 e “assembling” dos amplicões fragmentados, procedeu-se à validação e caracterização funcional dos genes correspondentes às enzimas alvo *in silico*.

2.1 Desenho dos primers “consenso” degenerados

No desenho dos primers “consenso” degenerados, utilizaram-se sequências em nucleótidos, presentes na base de dados ENA de EMBL-EBI, correspondentes às diferentes enzimas alvo (cutinase, feruloil esterase e lacase). Os alinhamentos múltiplos foram efetuados em separado e os primers desenhados de forma a emparelharem em regiões conservadas junto às extremidades dos genes, a fim de amplificarem sequências mais longas. Este procedimento encontra-se descrito detalhadamente na Secção 5.1 do capítulo II - Métodos, assim como, os primers “consenso” degenerados desenhados para as enzimas em estudo.

Inicialmente, no software MEGA converteram-se as sequências de nucleótidos para aminoácidos e realizaram-se alinhamentos múltiplos em aminoácidos através do software MUSCLE. A identificação de locais extremamente conservados foi efetuada com recurso à informação disponível na base de dados Pfam. As sequências que não continham as respetivas regiões conservadas foram excluídas dos alinhamentos.

As sequências foram novamente transformadas de aminoácidos para nucleótidos e os primers construídos em regiões conservadas de aproximadamente 20 nucleótidos consecutivos, o equivalente a 7 aminoácidos (Figura 25). Além disso, em cada primer foram incluídos no máximo 7 nucleótidos degenerados para cobrir os aminoácidos extremamente conservados de diferentes taxa bacterianas. No total, foram construídos

cinco primers para cutinases (Cut_F1, Cut_F2, Cut_R1, Cut_R2 e Cut_R3), quatro primers para feruloil esterases (Fae_F1, Fae_F2, Fae_R1 e Fae_R2) e quatro para lacases (Lac_F1, Lac_R1, Lac_R2 e Lac_R3).

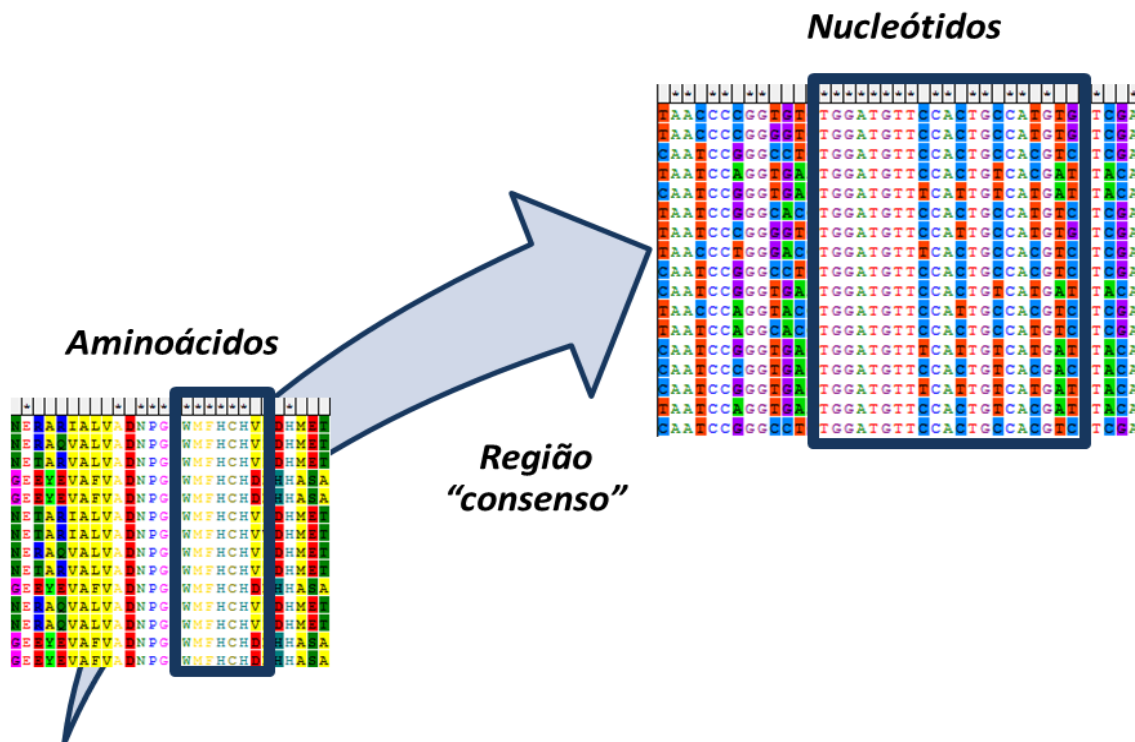


Figura 25 - Exemplo de uma região “consenso” de lacases identificada no software MEGA, através de um alinhamento múltiplo. O asterisco (*) no topo das sequências significa que para essa posição, o aminoácido ou nucleótido é comum a todas as sequências, refletindo a sua conservação.

As sequências de cutinases apresentavam diferenças no tamanho e nas zonas conservadas, sendo deste modo agrupadas em três tipos. Desta forma, para os três tipos de cutinases foram desenhados os seguintes primers: 1º) Cut_F1 e Cut_R1; 2º) Cut_F1 e Cut_R2; 3º) Cut_F2 e Cut_F3. Por este motivo, as amplificações por PCR de cutinases foram desdobradas em diferentes reações, onde se utilizaram diferentes pares de primers em separado para cada grupo de cutinases, em vez de, um “pool” de primers, na qual se utilizam em conjunto todos os primers na mesma reação, como foram os casos de feruloil esterases e lacases. Para uma melhor compreensão, na Figura 26 estão representados os primers utilizados nas 5 amplificações distintas, assim como o tamanho esperado dos fragmentos amplificados.

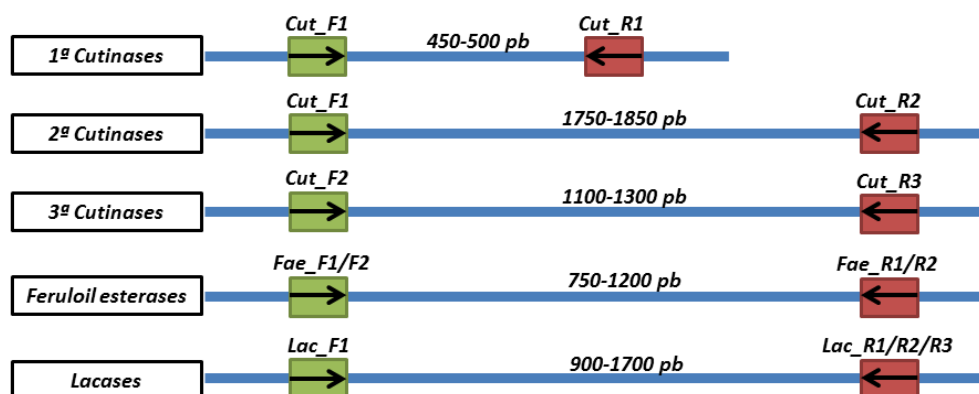


Figura 26 – Representação esquemática dos primers utilizados nas 5 amplificações previstas para as enzimas alvo, assim como o tamanho esperado nos amplicões.

2.2 Amplificação por PCR através de primers “consenso” degenerados

Na amplificação por PCR através de primers “consenso”, para as cutinases, as diferentes reações previstas foram desdobradas e, utilizado em separado apenas um par de primers por cada amplificação, mais propriamente, foram realizadas três amplificações com: Cut_F1 e Cut_R1; Cut_F1 e Cut_R2; e Cut_F2 e Cut_R3. Por sua vez, a metodologia utilizada para as feruloil esterases e lacases abrangeu uma “pool” de primers para cada uma das enzimas.

Conforme os alinhamentos múltiplos de sequências para as diferentes enzimas alvo, nas cinco amplificações por PCR otimizadas é esperada uma gama de valores para o tamanho dos fragmentos amplificados pelo par e/ou “pool” de primers em questão (Tabela 5). Na Figura 27 observam-se em gel de agarose os produtos de PCR obtidos nas diferentes amplificações: para as cutinases três reações em separado (um par de primers em cada) e para feruloil esterases e lacases uma reação (“pool” de primers). Desta forma, é possível comparar e confirmar os produtos de PCR amplificados pelos respectivos primers.

Tabela 5 – Gama de valores previstos para cada uma das cinco amplificações realizadas para as enzimas alvo.

Amplificação	Tamanho previsto (pb)
Cut F1-R1	450-500
Cut F1-R2	1750-1850
Cut F2-R3	1100-1300
Fae "Pool"	750-1200
Lac "Pool"	900-1700

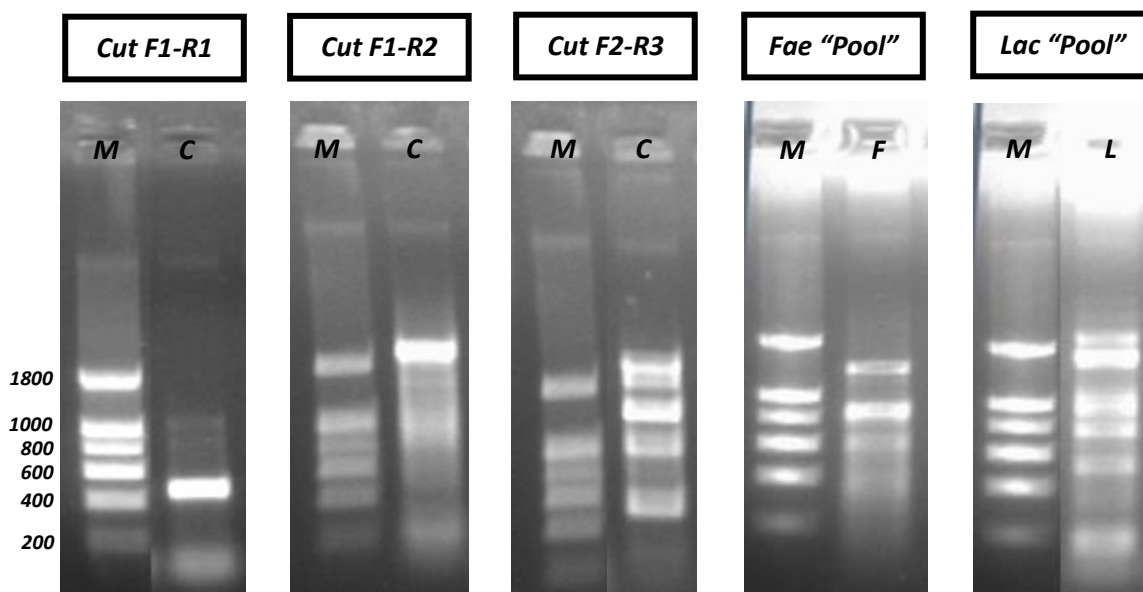


Figura 27 – Gel de agarose (1%) obtido após amplificação por PCR das diferentes enzimas alvo: C: produtos de PCR de cutinases; F: produtos de PCR de feruloil esterases; L: produtos de PCR de lacases. M: marcador NZYDNA Ladder I, com o tamanho das bandas representado do lado esquerdo.

Como se pode observar na Figura 27, a amplificação por PCR das classes distintas de enzimas alvo gerou amplicões de diferentes tamanhos, estando estes de acordo com os previstos, uma vez que as bandas de maior intensidade estão dentro da gama esperada (Tabela 6). Além disso, verificam-se algumas bandas fora dos valores esperados, o que provavelmente indica que ocorreu uma ligeira inespecificidade nas amplificações. Tal fato já seria de esperar, dado que os primers “consenso” apresentam em média 7 bases degeneradas num total de 20 pb. Contudo, após a sequenciação, os vários fragmentos amplificados são analisados e avaliados, mais concretamente, são relacionados com as respetivas enzimas, na qual são confirmados ou rejeitados *in silico*.

Nas diferentes amplificações, os amplicões gerados apresentam tamanhos superiores a 800 pb, com exceção na 1ª amplificação de cutinases (Cut F1-R1). Desta forma, estas amostras foram então fragmentadas por ultrasonificação, visto que devem ter tamanhos inferiores (entre 200 a 600 pb), requeridos na emulsão de PCR da pirosequenciação.

2.3 Fragmentação mecânica dos amplicões de enzimas alvo

Após a purificação dos produtos de PCR das várias amplificações descritas anteriormente, todas as amplificações foram sujeitas a fragmentação mecânica por ultrasonificação, com exceção da amplificação de cutinases Cut F1-R1, por apresentar fragmentos com um tamanho ideal para emulsão de PCR (200-600 pb).

Na Figura 28 estão representados os resultados obtidos para o perfil de fragmentação dos amplicões, apresentando um “smear” em cada uma das enzimas alvo, isto é, um contínuo de bandas pelo qual se pode observar os diferentes tamanhos dos amplicões fragmentados.

Os respectivos amplicões de cada uma das enzimas alvo foram fragmentados em separado, porém as duas amplificações correspondentes às cutinases, Cut F1-R2 e Cut F2-R3, foram misturadas no mesmo tubo na fragmentação mecânica (amostra C). Previamente ao processo de sequenciação, como pretendido para a emulsão de PCR são fragmentos entre 200 e 600 pb, pode verificar-se pela Figura 28 que o “smear” dos amplicões fragmentados vai desde 150 a 800 pb, no qual a sua grande maioria apresenta valores entre 200 e 600 pb, encontrando-se deste modo os fragmentos num tamanho adequado.

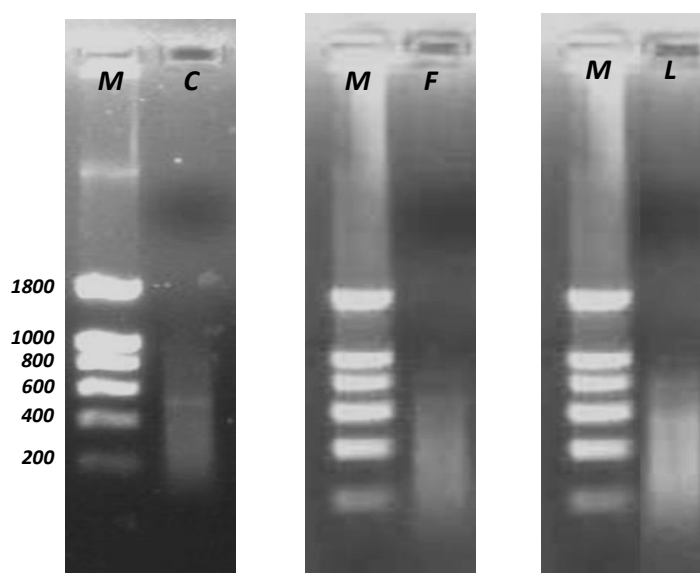


Figura 28 – Perfil de fragmentação das enzimas alvo obtido por ultrasonificação. Os fragmentos das enzimas alvo foram observados em gel de agarose (1%). Cutinases (C), feruloil esterases (F) e lacases (L). M: marcador NZYDNA Ladder I, cujo tamanho das bandas está representado do lado esquerdo.

3. Caraterização das enzimas alvo

3.1 Pirosequenciação dos amplicões gerados pelos primers “consenso”

As sequências obtidas pela sequenciação dos amplicões das enzimas alvo, na qual se utilizou a metodologia dos primers “consenso” degenerados, foram processadas para remoção de adaptadores e submetidas na “pipeline” de análise MG-RAST. A “pipeline” fornece uma atribuição filogenética e uma caraterização funcional de sequências de metagenomas, por comparação com bases de dados de proteínas e nucleótidos, apresentando também várias ferramentas para comparações de metagenomas.

Na Tabela 6 apresentam-se os resultados obtidos após o processamento de dados na pirosequenciação dos amplicões gerados pelos primers “consenso” degenerados. O número total de sequências geradas foi de 21950, na qual 90,7% passou no controlo de qualidade (19944 sequências de qualidade elevada). Por sua vez, o número total de pares de bases sequenciadas foi de 11880126, cujo comprimento médio das sequências apresentava 541 pares de bases. Mais uma vez, comprovou-se que a fragmentação mecânica por ultrasonificação foi apropriada para a sequenciação na plataforma 454 da Roche, visto que o comprimento médio das sequências está dentro dos valores requeridos na emulsão de PCR da pirosequenciação (200-600 pb).

Tabela 6 – Análise estatística do upload efetuado na “pipeline” de análise MG-RAST. Números e percentagem de sequências de qualidade elevada geradas no processamento de dados na pirosequenciação, assim como o número total de pares de bases e comprimento médio das sequências.

Amostra	# Total de sequências	# Sequências de qualidade elevada	% Sequências de qualidade elevada
Amplicões enzimas alvo	21950	19944	90,9
	# Total de pb	# Total de pb sequências de qualidade elevada	Comprimento médio sequências (pb)
	11880126	10892019	541

3.2 Identificação taxonômica das sequências amplificadas

Na Figura 29 está ilustrada a distribuição dos domínios, filo e classes taxonômicos correspondentes à anotação das sequências de qualidade elevada obtidas na pirosequenciação dos amplicões alvo. Esta informação foi prevista tendo em conta todas as anotações das bases de dados utilizadas pelo MG-RAST. Identifica-se portanto, uma elevada predominância do domínio *Eukaryota* com 93%, seguindo-se *Bacteria* com 5%, *Viruses* 2%, *Archaea* e outras sequências contabilizando um valor inferior a 1%.

Pode também observar-se que, 81% das sequências sequenciadas correspondem à classe *Insecta* e apenas 5% pertencem ao domínio *Bacteria*, o que significa que os primers “consenso” degenerados amplificaram sobretudo DNA hospedeiro – *Crematogaster scutellaris*.

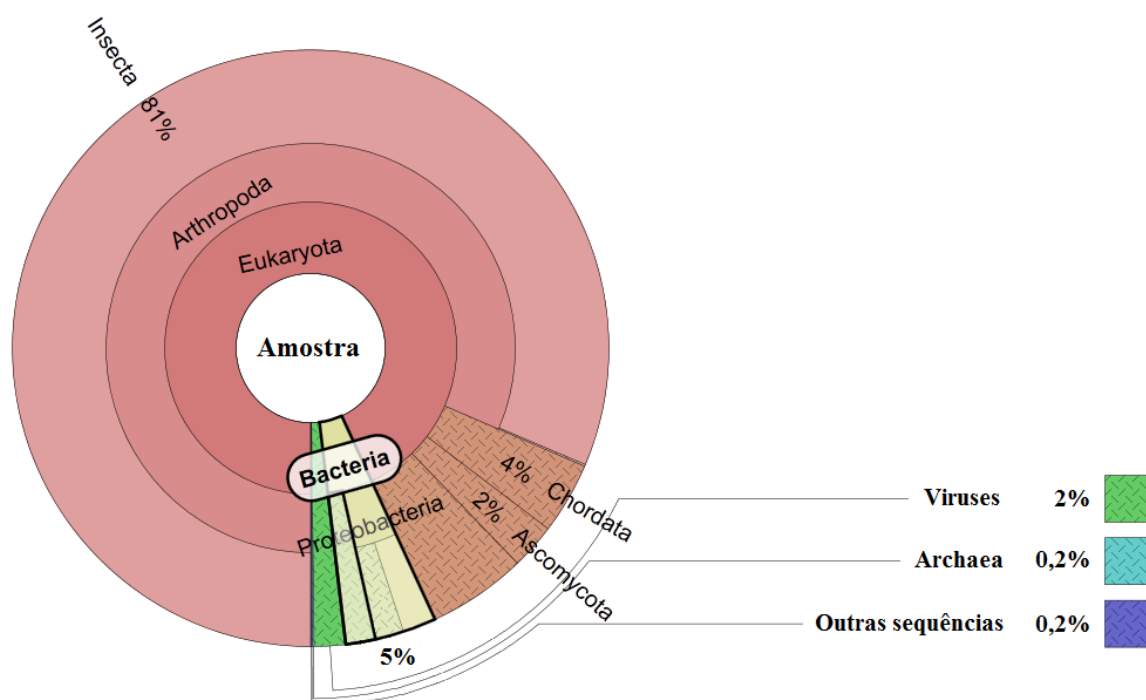


Figura 29 – Anotação taxonômica prevista pela “pipeline” de análise MG-RAST para as sequências de elevada qualidade amplificadas pelos primers “consenso” degenerados. A grande maioria das sequências amplificadas correspondem à classe *Insecta* (hospedeiro) e apenas 5% corresponde ao domínio *Bacteria*.

Tal facto não seria de esperar, uma vez que se pretendia amplificar regiões de genes bacterianos que codificassem as enzimas alvo cutinases, feruloil esterases e lacases. Deste modo, apesar do kit da QIAGEN utilizado na extração de DNA ser apropriado para bactérias, verificou-se que para amostras metagenómicas de regiões intestinais, na qual está presente uma grande parte de células eucariotas, ocorreu conjuntamente a lise celular de ambos (bactérias e células eucarióticas), o que conduziu a uma extração de DNA de células eucariotas e procariotas presentes na amostra. Para além disso, a estratégia de desenhar primers “consenso” degenerados para locais conservados de enzimas bacterianas é arriscada, uma vez que estes locais por vezes encontram-se também conservados nas enzimas de eucariotas.

Para além disso, como ainda não é conhecido nenhum genoma do género da formiga em estudo, *Crematogaster*, não foi possível verificar se os primers construídos emparelhavam com regiões do mesmo, sendo somente os seus “hits” testados em outros géneros de formigas presentes nas bases de dados.

Uma outra possibilidade dos primers “consenso” degenerados terem amplificado regiões inespecíficas, ou seja, DNA não pertencente a bactérias, poderá dever-se ao facto de as condições de amplificação terem sido otimizadas de acordo com o tamanho esperado e obtido nos fragmentos para as diferentes enzimas, cujo tamanho apenas foi confirmado pelas bandas observadas em gel de agarose e não por controlos positivos (Secção 2.2). Por exemplo, ao diminuir-se a temperatura de annealing das amplificações por PCR, verifica-se um aumento na intensidade da banda e na inespecificidade dos primers.

3.3 Identificação funcional das sequências amplificadas

A anotação funcional das sequências de qualidade elevada foi prevista pela “pipeline” de análise MG-RAST. Apesar de apenas 5% das sequências pertencerem ao domínio *Bacteria*, o que diminui substancialmente a possibilidade de as sequências corresponderem às enzimas alvo, foram encontradas 6 sequências codificantes de lacases bacterianas. Dentro destas, 5 foram anotadas como “copper resistance protein A” (CopA) e uma como “blue copper oxidase CueO precursor” (CueO), as quais pertencem todas à família de proteínas multi-cobre oxidases.

Tendo em conta a fragmentação dos amplicões por ultrasonificação, a procura de enzimas alvo foi ainda efetuada por uma estratégia de outro modo, diferente da anotação prevista pelo MG-RAST. Nesta medida, todas as sequências geradas na pirosequenciação dos amplicões alvo foram “assembladas” pelo GS Assembler da Roche, originando 255 contigs. Através do software Blastall, utilizando a ferramenta tBlastn, em todos os contigs foram procuradas regiões “consenso” em aminoácidos nas seis frames possíveis, mais propriamente, primers e locais conservados entre os primers para cada uma das enzimas alvo. Contudo, apesar de terem sido identificados primers e/ou regiões “consenso” de enzimas alvo nos contigs, as suas anotações, inferidas pela ferramenta Blastx do NCBI, apresentaram bastante homologia com regiões de genomas eucariotas, na qual todos os contigs foram rejeitados *in silico*.

Dando particularidade às lacases, dado que foi a única classe de enzima alvo identificada nas sequências de qualidade elevada, estas apresentam domínios altamente conservados nas regiões de ligação aos quatro átomos de cobre. Estas regiões estão organizadas ao longo da sequência de aminoácidos da seguinte forma: HXHG, HXH, HXXHXH e HCHXXHXXXXM/L, cujo X corresponde a qualquer aminoácido [129]. Recorrendo às bases de dados Pfam e InterPro pode verificar-se que a grande maioria das sequências codificantes de lacases apresentam três domínios, na qual se incluem as multicobre oxidases CopA e CueO, ordenados pelo seguinte modo: Cu-oxidase_3, Cu-oxidase_1 e Cu-oxidase_2.

Na Figura 30 estão representadas 4 sequências (A, B, C e D) amplificadas pelos primers “consenso” degenerados, que são fragmentos de lacases bacterianas. Para cada uma delas estão descritos os domínios conservados, assim como, as características do melhor hit identificados por Blastx na base de dados do NCBI. É de notar que, cada uma destas sequências foi anotada previamente pelas bases de dados utilizadas pela “pipeline” MG-RAST, de entre as quais se destacam Genbank, TrEMBL, RefSeq, eggNOG e PATRIC. Como referido anteriormente, no total das sequências de elevada qualidade anotaram-se 6 como lacases, contudo, duas correspondiam à sequência A e outras duas à sequência B, em frames diferentes. Desta forma, o número de sequências codificantes de lacases bacterianas foi reduzido para 4.

Relativamente à sequência A da Figura 30, a classificação do melhor hit corresponde a uma CopA com uma similaridade de 96% com a bactéria *Rhodococcus quingshengii*. Por sua vez, a sequência B corresponde também a uma CopA com 99% de semelhança com

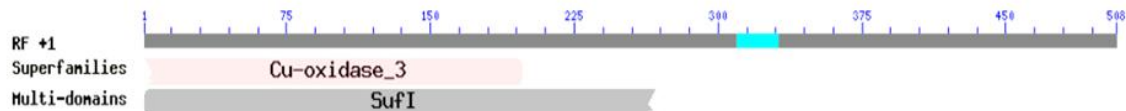
Rhodococcus erythropolis. Estas duas sequências correspondem provavelmente à mesma enzima do mesmo microrganismo (*Rhodococcus*), pois como se pode observar a sequência A apresenta como domínio conservado Cu-oxidase_3 e a sequência B o domínio Cu-oxidase_2, os quais estão presentes nas extremidades dos genes codificantes de lacases. Porém, como os produtos de PCR foram fragmentados aleatoriamente por ultrasonificação e, estatisticamente, como se amplificou muito mais DNA hospedeiro em relação ao bacteriano, poderá ter ocorrido “perda de informação” das sequências amplificadas, pois os fragmentos de sequências de tamanho inferior foram removidos na purificação com “beads” AMPure XP, previamente à sequenciação. Desta forma, não foi possível “assemblar” todas as sequências amplificadas pelos primers “consenso” degenerados.

A sequência C da Figura 30 foi identificada como CopA com 82% de similaridade com *Tetrasphaera elongata* e corresponde a uma zona entre as regiões altamente conservadas de ligação aos átomos de cobre – domínio Cu-oxidase_1. As “copper resistance protein A” (CopA) são lacases que normalmente apresentam translocação dependente de uma dupla arginina (TAT), característica observada num largo número de proteínas periplasmáticas que atravessam a membrana com metais ligados a co-factores. A combinação de ligações de átomos de cobre com TAT sugere mecanismos de resistência por empacotamento e exportação.

Por último, a sequência D da Figura 30 foi anotada como CueO com 89% de similaridade com a bactéria *Serratia proteamaculans*, cuja região descrita corresponde ao domínio Cu-oxidase_1. A enzima CueO exibe uma atividade oxidativa numa vasta variedade de substratos fenólicos e, foi inicialmente descrita em *Escherichia coli* como uma lacase envolvida na resistência ao cobre [130]. O domínio designado por PRK10965, identificado pela base de dados NCBI, é característico da família de proteínas multi-cobre oxidases.

Os géneros *Serratia* e *Rhodococcus* foram identificados pela região V3-V4 dos genes do rRNA 16S e várias sequências destes foram incluídas nos alinhamentos para a construção de primers, o que revela coerência na anotação das sequências A, B e D. A sequência C foi a que apresentou menor homologia (82%) perante a base de dados do NCBI, na qual a sua taxonomia *Tetrasphaera elongata* não foi previamente identificada pelos genes do rRNA 16S. Tais resultados não descartam a possibilidade desta sequência ser uma multi-cobre oxidase putativa, pois o género *Tetrasphaera* pertence à ordem *Actinomycetales*, da qual também faz parte o género *Rhodococcus*.

Sequência A

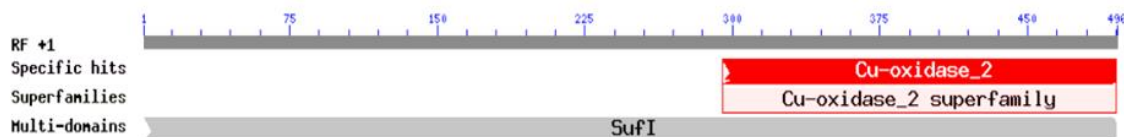


Copper resistance protein A [*Rhodococcus quingshengii*]

Range 1: 122 to 289

Score	Expect	Method	Identities	Positives	Gaps	Frame
311 bits(797)	3e-101	Compositional matrix adjust.	161/168(96%)	165/168(98%)	0/168(0%)	+1
Query 1	TTVHWHG	I A V R N D M D G V P P V T Q E R I R P G A E F T Y D F V A P D H G T Y W H S H S G L Q A D R G L F G A	180			
Sbjct 122	TSVHWHG	I A R N D M D G V P P V T Q E R I Q P G A E F T Y D F V A P D H G T Y W H S H S G L Q A D R G L F G A	181			
Query 181	LVVEDPNDRTGADEDVVIVIDDWLDGLGTTTPDAVLRALSPDVQgghhghgg	T P P E V D G A E	360			
Sbjct 182	LVVEDNDRTGADEDVVIVIDDWLDGLGTTTPDAVLRALSPDVQGGHHGHGGT	P P E V D G A E	241			
Query 361	IDSAAELLGSGHGSNRALGGTATHITYPLHLINGWPFVEDPFVVSATAG	504				
Sbjct 242	IDSAAELLGSGHGSNRALGGTATHITYPLHLINGWPFVEDPFVVSATAG	289				

Sequência B

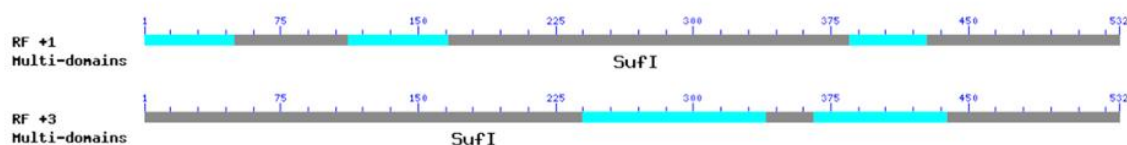


Copper resistance protein A [*Rhodococcus erythropolis*]

Range 1: 338 to 502

Score	Expect	Method	Identities	Positives	Gaps	Frame
340 bits(872)	4e-116	Compositional matrix adjust.	163/165(99%)	164/165(99%)	0/165(0%)	+1
Query 1	QRYDVLVTVQSGSWPIAAKVEGKSGGVACLLRTPDSVAVVDLTSPAVLSGSGGRLVLESE	180				
Sbjct 338	QRYDVLVTVQSGSWPIAAKVEGKSGGVACLLRTPDSVAVVDLTSPAVLSGSGGRLVLESE	397				
Query 181	LRPSGALEQRAPDRAYSVDLIQAGDRYLWGMAGADAGRLRMKQGERIRIVMNNNDTDMVHP	360				
Sbjct 398	LRPSGALEQRAPDRAYSVDLIQAGDRYLWGMAGADAGRLRMKQGERIRIVMNNNDTDMVHP	457				
Query 361	MHTHCHTFSVPSYGGLRRTVIALPRTQMAIDFDADNPGRWMEHC	495				
Sbjct 458	MHTHCHTFSVPSYGGLRRTVIALPRTQMAIDFDADNPGRWMEHC	502				

Sequência C



Copper resistance protein A [*Tetrasphaera elongata*]

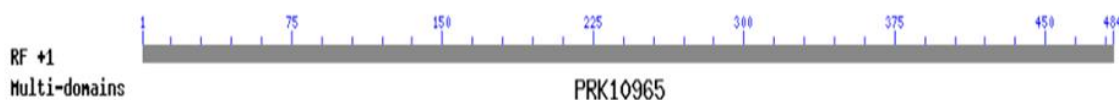
Range 1: 303 to 409

Score	Expect	Method	Identities	Positives	Gaps	Frame
149 bits(377)	2e-54	Compositional matrix adjust.	88/107(82%)	95/107(88%)	0/107(0%)	+1
Query 208	LVGTGFTITHSDGHAVEPTEVGAFYIGMGERYDAIVTLADGAPFLVARPFPGKTSGGQAMa					387
Sbjct 303	L G TITHSDGHAVEPTEVGAFYIGMGERYDAIVTL DG FPLVARPFPGKTSGGQA+A					362
Query 388	vvrtgagaappvaaDPSELAGQVLIGSQLRPTEAAKLPGEKPDATVD			528		
Sbjct 363	VVRTG+G+AP V A P+EL+GQVLIGSQLRP E+AKLP + PDATVD					409

Range 2: 254 to 308

Score	Expect	Method	Identities	Positives	Gaps	Frame
88.2 bits(217)	2e-54	Compositional matrix adjust.	42/55(76%)	44/55(80%)	0/55(0%)	+3
Query 60	WGDAGDVITYPLFLINKPPTTRTCWPPSRGQKVRVINAASDTIFTVALGGHRL			224		
Sbjct 254	WGDAGDVITYPLFLINKPPTTRTCWPPSRGQKVRVINAASDTIFTVALGGHRL				308	

Sequência D



Blue copper oxidase CueO [*Serratia proteamaculans*]

Range 1: 150 to 308

Score	Expect	Method	Identities	Positives	Gaps	Frame
281 bits(718)	2e-89	Compositional matrix adjust.	141/159(89%)	142/159(89%)	0/159(0%)	+1
Query 1	GQQVMKGLAGLVLEDDDESSKLPKPTWGQDDIPVILQDKRLGKDAQIEYQLDVMSAAVG			180		
Sbjct 150	GQQVMKGLAGLVLEDDDESSKLPKPTWGQDDIPVILQDKRLGKDAQIEYQLDVMSAAVG				209	
Query 181	GSATVCSPTVLTYPYPRHVA PRGWLRRLRFLNGCNARSINLAASDNRP LYVIASDGGFLAEPV			360		
Sbjct 210	WFGDRMFTNGAQYPRHVA PRGWLRRLRFLNGCNARSINLATSDNRP LYVIASDGGFLAEPV				269	
Query 361	KLTELMLMGERFEVLVDASDGKAFDIVTLFVKQMG L			477		
Sbjct 270	KLTELMLMGERFEVLVDASDGKAFDIVTLFVKQMG L				308	

Figura 30 – Anotação das sequências correspondentes a lacases. Paracada uma delas estão descritos os domínios conservados, assim como, as características do melhor hit identificado por Blastx na base de dados do NCBI. A sequência A, B e C correspondem a CopA de *Rhodococcus quingshengii*, *Rhodococcus erythropolis* e *Tetrasphaera elongata*, respetivamente, enquanto a sequência D corresponde a CueO de *Serratia proteamaculans*. Nas sequências A e B, as regiões altamente conservadas de ligação aos átomos de cobre estão assinaladas a azul, mais propriamente, HXHG e HXH na sequência A e, HXXHXH e a parte inicial de HCHXXXHXXXM/L na sequência B.

Na Figura 31 estão representadas quatro árvores filogenéticas, correspondentes às sequências A, B, C e D anotadas como lacases (CopA e CueO). Foram ainda incluídas sequências de gêneros identificados pelos genes dos rRNA 16S, utilizadas na construção dos primers “consenso” degenerados, assim como, sequências de espécies com elevada homologia identificadas pelos “hits” do Blastx do NCBI.

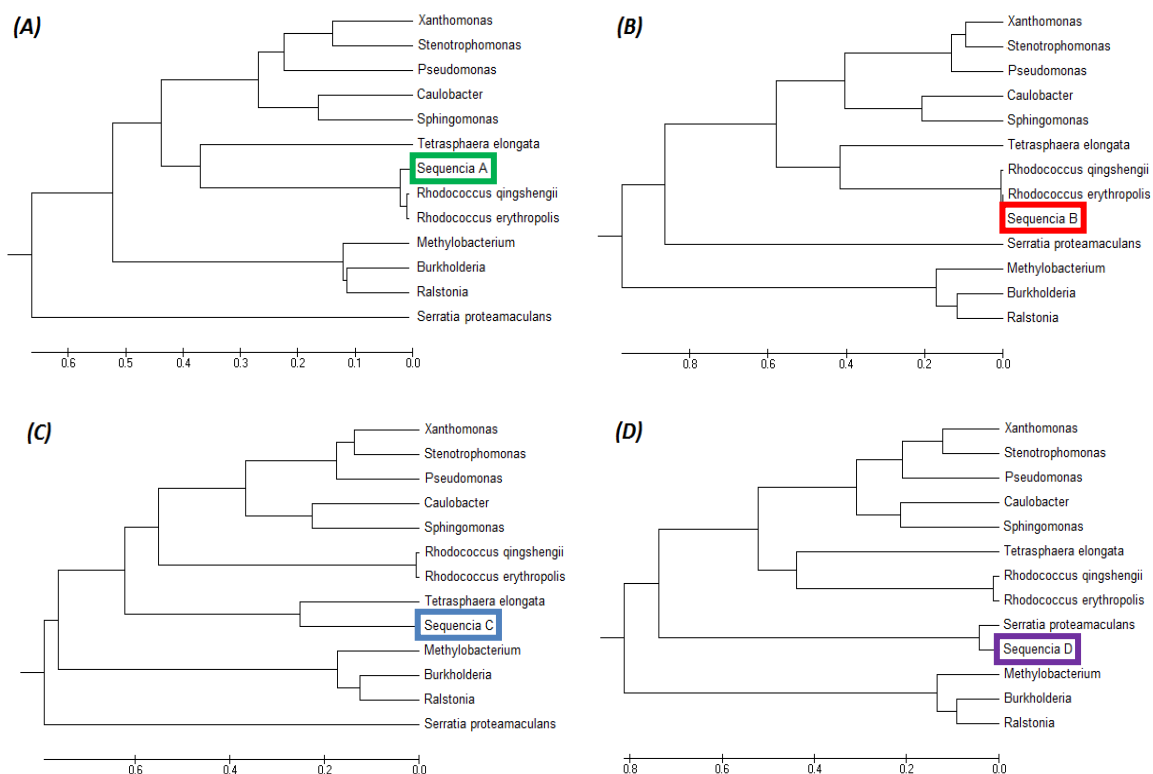


Figura 31 – Representação de quatro árvores filogenéticas obtidas no software MEGA, correspondentes às sequências A, B, C e D anotadas como lacases (CopA e CueO). As árvores filogenéticas incluem sequências de gêneros identificados pelos genes dos rRNA 16S, utilizadas na construção dos primers “consenso” degenerados e sequências com elevada homologia, identificadas pelos “hits” do Blastx do NCBI.

Apesar das sequências A e B apresentarem elevada homologia com o gênero *Rhodococcus*, 96% e 99% respectivamente, a espécie difere para cada uma delas, pois a sequência A corresponde à espécie *Rhodococcus qingshengii* e a sequência B a *Rhodococcus erythropolis*. Além disso, junto aos locais conservados de ligação aos átomos de cobre (Figura 30), evidenciam-se algumas disparidades nas duas sequências a nível de aminoácidos, o que poderá significar que estamos perante uma enzima com características diferentes das já conhecidas.

O género *Rhodococcus* apresenta um vasto número de atividades enzimáticas, uma estrutura de parede celular única e propriedades adequadas para vários processos biotecnológicos, como biotransformação e degradação de vários compostos aromáticos [131]. Este género foi já isolado de vários habitats, como por exemplo solo (*Rhodococcus quingshengii*), água do mar (*Rhodococcus erythropolis*) e intestino de térmitas (*Rhodococcus opacus*) [132; 133]. Por sua vez, *Serratia* é caracterizado como um patógeno oportunista, destacando-se pela sua pigmentação e infeções nosocomiais em humanos [134]. Encontra-se distribuído em diversos ambientes como água, solo e também em intestinos de animais roedores.

Na sequência D verifica-se uma homologia de 89% com o género *Serratia*, a qual apresenta uma diferença de 12 aminoácidos consecutivos, podendo concluir-se que se trata de uma nova enzima. Desta forma, seria interessante fazer a identificação do gene completo, assim como a sua clonagem em vetores recombinantes, uma vez que esta poderá exibir atividades enzimáticas específicas em substratos fenólicos presentes na lenhina de cortiça.

Enquanto que a lacase Suberase® (Novozymes), enzima atualmente utilizada na indústria corticeira, é proveniente de fungos, as enzimas anotadas neste estudo são provenientes de bactérias e, no entanto, apresentam as mesmas potencialidades na deslenhificação da cortiça. Nesta medida, poderá ser possível uma eventual substituição de lacases existentes, uma vez que estas lacases bacterianas poderão apresentar propriedades melhoradas.

Na Figura 32 ilustra-se um PSI-Blast efetuado no Swiss Institute of Bioinformatics (SIB) para cada uma das sequências, no qual se tem uma melhor perceção da localização da sequência anotada em relação à totalidade do gene homólogo. Na parte superior da imagem encontra-se a localização das sequências A, B e C em relação ao gene homólogo “copper resistance protein A”, enquanto que na parte inferior apresenta-se a localização da sequência D em relação ao gene “blue copper oxidase CueO”, mais propriamente, correspondente ao domínio intermédio Cu-Oxidase_1.

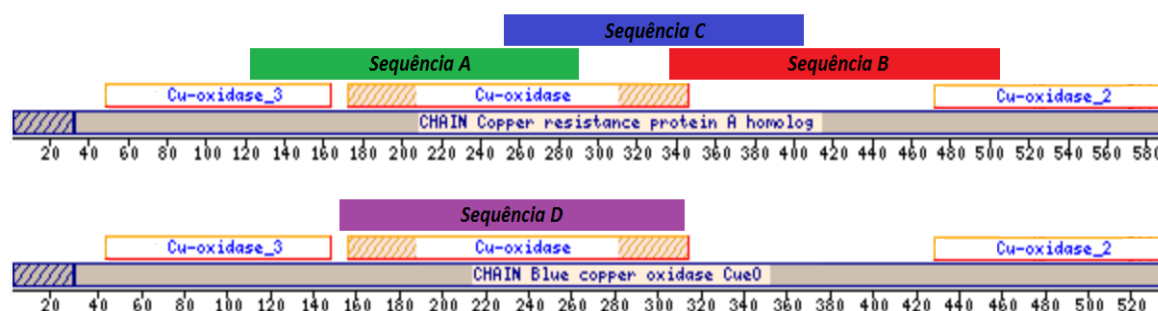


Figura 32 – Representação do PSI-Blast realizado no Swiss Institute of Bioinformatics (SIB) para cada uma das sequências parciais anotadas (A, B, C e D), a fim de identificar-se a que região do gene homólogo corresponde.

Com este estudo, evidencia-se a oxidação de lenhina, presente na cortiça, pelo microbioma intestinal da formiga, dado que as bactérias apresentam no seu genoma genes codificantes de lacases (CopA e CueO). Contudo, as sequências anotadas são de regiões parciais dos genes codificantes de lacases, sendo necessário efetuar estudos posteriores para se obter a sequência completa dos genes e testar a sua atividade enzimática perante diversos substratos fenólicos. É de notar que, a única enzima utilizada pela indústria corticeira no tratamento da cortiça é uma lacase (Suberase®, Novozymes), tornando o presente estudo importante para esta área de investigação, podendo ser um chamariz a novas aplicações industriais.

O conhecimento e elucidação de outras suberinas e vias metabólicas envolvidas na digestão e bioconversão de cortiça, presentes nos tecidos intestinais da formiga, poderá ser desvendado por intermédio de abordagem metagenómica “whole genome shotgun” (WGS), na qual é sequenciado e analisado todo o DNA, ou seja, tanto do hospedeiro (formiga) como das comunidades microbianas, resultando num perfil de todos os genes presentes na amostra metagenómica.

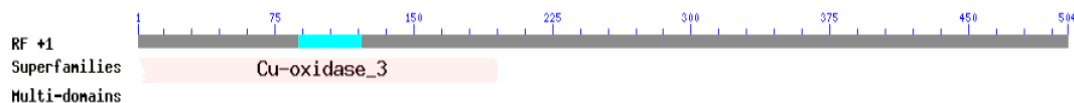
3.3.1 Identificação de lacases eucariotas

Apesar das sequências amplificadas corresponderem maioritariamente ao domínio *Eukaryota*, entre estas foram também previstas duas lacases correspondentes a genes eucariotas, pela “pipeline” de análise MG-RAST. Tal facto deve-se por os genes codificantes de lacases apresentarem, em todos os domínios, regiões bastante conservadas na ligação dos átomos de cobre, local onde os primers foram desenhados.

Na Figura 33 estão representadas 2 sequências (E e F) amplificadas pelos primers “consenso” degenerados, correspondentes aos fragmentos de lacases eucariotas. Para cada sequência estão descritos os domínios conservados, assim como, as características do melhor hit identificados por Blastx na base de dados do NCBI. As sequências foram anotadas como “diphenol oxidase” (sequência E) e “L-ascorbate oxidase” (sequência F). A primeira apresenta uma similaridade de 53% com a espécie *Cryptococcus gattii* (domínio *Fungi*), e a segunda uma homologia de 74% com a espécie *Camponotus floridanus* (formiga “carpinteira” que nidifica na madeira). Ambas as sequências correspondem ao primeiro domínio da família multi-cobre oxidase (Cu-oxidase_3).

Nesta medida, podemos inferir que a oxidação de lenhina da cortiça é provavelmente efetuada por lacases exógenas e endógenas, mais propriamente, enzimas presentes no hospedeiro (origem endógena) e nas bactérias e fungos (origem exógena).

Sequência E

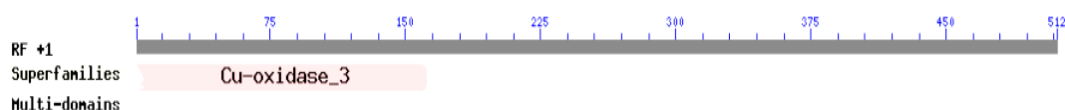


Diphenol oxidase [*Cryptococcus gattii*]

Range 1: 104 to 248

Score	Expect	Method	Identities	Positives	Gaps	Frame
147 bits(371)	2e-38	Compositional matrix adjust.	80/151(53%)	99/151(65%)	6/151(3%)	+1
Query 4	T HWHG L NQNGTQAQDGVPAFTQCP	IPAggtytytftvtgQWGTYW	HSHTAVQYTDGLY	183		
Sbjct 104	S HWHG L RQKDTFYM	DGVP TQCP	IPGGSYTYNFTISDQSGTYW	HSHTAVQYTDGLY	163	
Query 184	GPMVFHSAKEKYQRGRDYDEEVLLFFSDWYHDRGETIANALNSVQGYKLTNAAPPSQSVL	GP++ HS E QRGRDYDE++F SDW HD E I AL + GY+ + A P S+L		363		
Sbjct 164	GPIVHVSVEPIQRGRDYDEDRIVFVSDWLHDDSEIIIAALATPAGYRGSPAPFQGSIL			223		
Query 364	INGRGQYNCSNVSKSTDPSSYSCITPAWPEL		456			
Sbjct 224	INGRGQ TNC-----TATNSSSCYYPAPPEI		248			

Sequência F



L-ascorbate [*Camponotus floridanus*]

Range 1: 161 to 225

Score	Expect	Method	Identities	Positives	Gaps	Frame
113 bits(282)	7e-26	Composition-based stats.	48/65(74%)	53/65(81%)	0/65(0%)	+1
Query 1	T S HWHG L HHLTPYMDGVPYVSQCPIHPGATFQYNYIATEAGTHFWHSH	T GMLNTFQIF	180			
Sbjct 161	T T HWHG L HHLTPYMDGVPYVSQCPI PGATF+YNYIA+EAGTHFWHSH	G +F	220			
Query 181	AFDIV	195				
Sbjct 221	GSLIV	225				

Range 2: 206 to 269

Score	Expect	Method	Identities	Positives	Gaps	Frame
109 bits(272)	1e-24	Composition-based stats.	47/71(66%)	58/71(81%)	7/71(9%)	+1
Query 298	FYHCNVCTVYLSAGFQRGDGVFGPLIVRVPPKINWVKDLYDHEYTLVVTDWIHELSTSK	F+H ++ GFQRGDGVFG LIVRVPPK+NWHKDLYD+DE+ LVV DW+HEL +K	477			
Sbjct 206	FWHSHI-----GFQRGDGVFGSLIVRVPPKVNWVKDLYDDEHILVVADWVHELGLTK		258			
Query 478	FLAHYHADGNN	510				
Sbjct 259	FLAHHSDGHN	269				

Figura 33 – Sequências E e F anotadas pela “pipeline” MG-RAST correspondes a lacases eucariotas. Para cada uma apresentam-se os domínios conservados, assim como, as características do melhor hit identificado por Blastx na base de dados do NCBI. A sequência E foi anotada como “diphenol oxidase” correspondente a *Cryptococcus gattii* e a sequência F como “L-ascorbate oxidase” de *Camponotus floridanus*. As regiões altamente conservadas de ligação aos átomos de cobre estão assinaladas a azul , mais propriamente, HXHG e HXH.

V - CONCLUSÕES

O objetivo de estudo deste trabalho consistiu na caracterização de novas suberinases bacterianas, presentes no microbioma intestinal de formigas da espécie *Crematogaster scutellaris*. Dado que as formigas em questão nidificam preferencialmente na cortiça de sobreiros, acredita-se que têm a capacidade de degradar a suberina, principal constituinte da parede celular nas células da cortiça.

Para tal, foi estabelecida uma abordagem metagenômica que consistiu na identificação das bactérias presentes no microbioma intestinal da formiga, para facilitar a seleção de genes correspondentes às enzimas alvo (cutinases, feruloil esterases e lacases) e a construção de primers para regiões consenso presentes nas diferentes classes destas enzimas. Os genes alvo foram amplificados pelos primers “consenso” e pirosequenciados a partir do DNA da comunidade bacteriana.

Esta estratégia permitiu identificar a estrutura e composição da comunidade microbiana do gáster da formiga, revelando uma elevada diversidade e uma predominância bacteriana no filo *Proteobacteria* (68,0%), seguindo-se os filos *Bacteroidetes* (17,2%), *Firmicutes* (7,0%), *Acidobacteria* (4,6%) e *Actinobacteria* (3,2%). Na presente comunidade foram identificados microrganismos com enzimas alvo já descritas em bases de dados, nomeadamente ENA e UniProt, que serviram para a construção dos primers “consenso” degenerados.

A estratégia funcionou bem para as lacases, identificando-se 4 sequências bacterianas pertencentes à família multi-cobre oxidase, sendo 3 anotadas como “copper resistance protein A” e uma como “blue copper oxidase CueO precursor”. As CopA apresentaram elevada homologia com o género *Rhodococcus* e *Tetrasphaera* e, a CueO com o género *Serratia*, podendo então concluir-se que possivelmente se tratam de novas enzimas, com capacidade de oxidação de compostos fenólicos presentes na lenhina da cortiça. No entanto não foi possível identificar sequências para as restantes enzimas alvo (cutinases e feruloil esterases).

Além disso, foram ainda anotadas 2 sequências de lacases eucariotas, “diphenol oxidase” e “L-ascorbate oxidase”, pertencente ao fungo *Cryptococcus* e à formiga “carpinteira” *Camponotus floridanus*, respetivamente. Estas sequências apresentam a mesma organização de 3 domínios, tais como as lacases bacterianas, inferindo-se também a possibilidade de oxidação de lenhina por parte de enzimas presentes no hospedeiro (*Crematogaster scutellaris*) e fungos.

A estratégia estabelecida não foi completamente sucedida, o que revela a dificuldade na identificação de alguns genes parciais através de sequências em aminoácidos. Uma estratégia alternativa poderá ser a análise total de DNA presente na amostra. Devido ao tamanho reduzido do gáster da formiga e à contaminação do DNA com células da própria formiga, esta estratégia apresenta também algumas fragilidades.

Apesar da indicação da presença de lacases bacterianas no microbioma intestinal da formiga serão necessários vários estudos para desvendar a totalidade dos genes codificantes de lacases, ou seja, como as sequências anotadas são de regiões parciais de genes codificantes de lacases, será fundamental a obtenção das sequências completas dos mesmos. Desta forma, futuramente poderá ser realizada a clonagem destes genes bacterianos em vetores recombinantes e verificada a sua especificidade e atividade enzimática perante vários substratos fenólicos. Por outro lado, a sequenciação do metagenoma do gáster da formiga e/ou a construção de outros conjuntos de primers menos degenerados (mais específicos) permitirão a identificação de cutinases e feruloil esterases.

Em suma, Portugal sendo o maior produtor de cortiça do mundo faz com que esta matéria-prima seja tão relevante para a economia nacional, neste estudo prova-se que a descoberta de novas lacases é bastante interessante para ajudar no melhoramento da qualidade da cortiça. Deste modo, esta investigação aponta também para novas suberinases, com capacidade de bioconversão da cortiça em novos compostos de valor acrescentado, sendo assim muito importante reduzir os deperdícios desta magnífica matéria-prima, podendo estes ser aplicados em variadas indústrias.

VI - BIBLIOGRAFIA

- [1]APCOR, Cultura, natureza, futuro: Cortiça em números. Cork Information Bureau (2010).
- [2]<http://www.oje.pt/noticias/destaque/portugal-domina-exportacoes-mundiais-de-cortica-lideradas-pela-corticeira-amorim>. (Acedido a 2 de Janeiro de 2013).
- [3]J.S. Pereira, M.N. Bugalho, M.C. Caldeira, Do sobreiro à cortiça: Um sistema sustentável. (2008).
- [4]N. Cordeiro, M.N. Belgacem, A. Gandini, C.P. Neto, Urethanes and polyurethanes from suberin 2: synthesis and characterization. *Industrial Crops and Products* 10 (1999) 1-10.
- [5]N. Cordeiro, M.N. Belgacem, A.J.D. Silvestre, C.P. Neto, A. Gandini, Cork suberin as a new source of chemicals. 1. Isolation and chemical characterization of its composition. *International Journal of Biological Macromolecules* 22 (1998) 71-80.
- [6]N. Cordeiro, Fraccionamento da cortiça e caracterização dos seus componentes. Estudo de possibilidades e valorização da cortiça. Tese Doutoramento - Universidade de Aveiro (1998).
- [7]H. Pereira, Chemical composition and variability of cork from *Quercus suber L.* *Wood Science and Technology* 22 (1988) 211-218.
- [8]C. Coquet, E. Ferre, D. Peyronel, C. Dal Farra, A.M. Farnet, Identification of new molecules extracted from *Quercus suber L.* cork. *Comptes Rendus Biologies* 331 (2008) 853-858.
- [9]J. Graca, S. Santos, Suberin: a biopolyester of plants' skin. *Macromolecular Bioscience* 7 (2007) 128-135.
- [10]S.M. Rocha, B.J. Goodfellow, I. Delgadillo, C.P. Neto, A.M. Gil, Enzymatic isolation and structural characterisation of polymeric suberin of cork from *Quercus suber L.* *International Journal of Biological Macromolecules* 28 (2001) 107-119.
- [11]N. Cordeiro, N.M. Belgacem, A. Gandini, C.P. Neto, Cork suberin as a new source of chemicals: 2. Crystallinity, thermal and rheological properties. *Bioresource Technology* 63 (1998) 153-158.
- [12]M.A. Bernards, N.G. Lewis, The macromolecular aromatic domain in suberized tissue: a changing paradigm. *Phytochemistry* 47 (1998) 915-933.
- [13]R. Jarvinen, Cuticular and suberin polymers of edible plants: analysis by gas chromatographic-mass spectrometric and solid state spectroscopic methods. University of Turku (2010).

- [14]J. Graca, Hydroxycinnamates in suberin formation. *Phytochemistry Reviews* 9 (2010) 85-91.
- [15]S.R. Programme, Development of novel biotools for modification of bark components. Wood Material Science and Engineering Research Programme (2007).
- [16]H. Kontkanen, A. Westerholm-Parvinen, M. Saloheimo, M. Bailey, M. Ratto, I. Mattila, M. Mohsina, N. Kalkkinen, T. Nakari-Setälä, J. Buchert, Novel *Coprinopsis cinerea* polyesterase that hydrolyzes cutin and suberin. *Applied and Environmental Microbiology* 75 (2009) 2148-2157.
- [17]S.I. Vitorino, E.S.G. Neves, F. Gaspar, J.J.F. Marques, M.V.S. Romão, Suberin utilization by *Chrysonilia sitophila*: Evidence for lipolytic enzymes production. *Ciência e Técnica Vitivinícola* 22 (2007) 1-4.
- [18]C. Carvalho, M. Barros, J. Cabral, Cutinase structure, function and biocatalytic applications. *Electronic Journal of Biotechnology* 1 (1998) 1-14.
- [19]K. Dutta, S. Sen, V. Veeranki, Production, characterization and applications of microbial cutinases. *Process Biochemistry* 44 (2009) 127-134.
- [20]C.M. Carvalho, M.R. Aires-Barros, J.M. Cabral, Cutinase: from molecular level to bioprocess development. *Biotechnology and Bioengineering* 66 (1999) 17-34.
- [21]D.W. Wong, Feruloyl esterase: a key enzyme in biomass degradation. *Applied Biochemistry and Biotechnology* 133 (2006) 87-112.
- [22]A.E. Fazary, Y.H. Ju, Feruloyl esterases as biotechnological tools: current and future perspectives. *Acta Biochimica et Biophysica Sinica* 39 (2007) 811-828.
- [23]T. Koseki, S. Fushinobu, Ardiansyah, H. Shirakawa, M. Komai, Occurrence, properties, and applications of feruloyl esterases. *Applied Microbiology and Biotechnology* 84 (2009) 803-810.
- [24]E. Topakas, C. Vafiadi, P. Christakopoulos, Microbial production, characterization and applications of feruloyl esterases. *Process Biochemistry* 42 (2007) 497-509.
- [25]V. Madhavi, S.S. Lele, Laccase: Properties and Applications. *Bioresources* 4 (2009) 1694-1717.
- [26]P. Sharma, R. Goel, N. Capalash, Bacterial laccases. *World Journal of Microbiology and Biotechnology* 23 (2007) 823-832.
- [27]C. Galhaup, S. Goller, C.K. Peterbauer, J. Strauss, D. Haltrich, Characterization of the major laccase isoenzyme from *Trametes pubescens* and regulation of its synthesis by metal ions. *Society for General Microbiology* 148 (2002) 2159-2169.

- [28]A.J. Augustine, C. Kjaergaard, M. Qayyum, L. Ziegler, D.J. Kosman, K.O. Hodgson, B. Hedman, E.I. Solomon, Systematic perturbation of the trinuclear copper cluster in the multicopper oxidases: the role of active site asymmetry in its reduction of O₂ to H₂O. *Journal of the American Chemical Society* 132 (2010) 6057-6067.
- [29]D.M. Ivnitski, C. Khripin, H.R. Luckarift, G.R. Johnson, P. Atanassov, Surface characterization and direct bioelectrocatalysis of multicopper oxidases. *Electrochim Acta* 55 (2010) 7385-7393.
- [30]Shraddha, R. Shekher, S. Sehgal, M. Kamthania, A. Kumar, Laccase: microbial sources, production, purification, and potential biotechnological applications. *Enzyme Research* 2011 (2011) 217861.
- [31]A.M. Mayer, R.C. Staples, Laccase: new functions for an old enzyme. *Phytochemistry* 60 (2002) 551-565.
- [32]T.D.H. Bugg, M. Ahmad, E.M. Hardiman, R. Rahmanpour, Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports* 28 (2011) 1883-1896.
- [33]L. Ausec, J.D. van Elsas, I. Mandic-Mulec, Two- and three-domain bacterial laccase-like genes are present in drained peat soils. *Soil Biology & Biochemistry* 43 (2011) 975-983.
- [34]L. Ausec, M. Zakrzewski, A. Goesmann, A. Schluter, I. Mandic-Mulec, Bioinformatic Analysis Reveals High Diversity of Bacterial Genes for Laccase-Like Enzymes. *Plos One* 6 (2011).
- [35]T.D.H. Bugg, M. Ahmad, E.M. Hardiman, R. Singh, The emerging role for bacteria in lignin degradation and bio-product formation. *Current Opinion in Biotechnology* 22 (2011) 394-400.
- [36]K. Koschorreck, S.M. Richter, A.B. Ene, E. Roduner, R.D. Schmid, V.B. Urlacher, Cloning and characterization of a new laccase from *Bacillus licheniformis* catalyzing dimerization of phenolic acids. *Applied Microbiology and Biotechnology* 79 (2008) 217-224.
- [37]J.D. Majer, Ants: Bio-Indicators of Minesite Rehabilitation, Land-Use and Land Conservation. *Environmental Management* 7 (1983) 375-383.
- [38]M. Paiva, M.J. Way, M. Cammel, Estudo preliminar sobre a distribuição das formigas nos sistemas florestais em Portugal. *Boletim da Sociedade Portuguesa de Entomologia* 4 (1990) 197-208.

- [39]F. Soria, M. Villagrán, M. Ocete, Estudio del comportamiento alimentario de *Crematogaster scutellaris* Oliv. (Hym. Formicidae) en tres alcornocales del SW español. Boletín de Sanidad Vegetal Plagas 20 (1994) 637-642.
- [40]C. Collingwood, A. Prince, A guide to ants of continental Portugal (*Hymenoptera:Formicidae*). Sociedade Portuguesa de Entomologia 5 (1998).
- [41]M. Villagrán, Estudios biológico sobre *Crematogaster scutellaris* Olivier (*Hymenoptera, Formicidae*) en alcornocales de Andalucía occidental. Universidad de Sevilla Tese de Doutoramento (1992).
- [42]<http://www.hormigas.org/xEspecies/Crematogaster%20scutellaris.htm>. (Acedido a 2 de Janeiro de 2013).
- [43]S.M. Geib, T.R. Filley, P.G. Hatcher, K. Hoover, J.E. Carlson, M. Jimenez-Gasco Mdel, A. Nakagawa-Izumi, R.L. Sleighter, M. Tien, Lignin degradation in wood-feeding insects. Proceedings of the National Academy of Sciences of the United States of America 105 (2008) 12932-12937.
- [44]M.E. Scharf, A. Tartar, Termite digestomes as sources for novel lignocellulases. Biofuels Bioproducts and Biorefining 2 (2008) 540-552.
- [45]Z. Mendel, A. Protasov, M. Sharon, A. Zveibil, S. Ben Yehuda, K. O'Donnell, R. Rabaglia, M. Wysoki, S. Freeman, An Asian ambrosia beetle *Euwallacea fornicatus* and its novel symbiotic fungus *Fusarium sp* pose a serious threat to the Israeli avocado industry. Phytoparasitica 40 (2012) 235-238.
- [46]M.E. Scharf, D.G. Boucias, Potential of termite-based biomass pre-treatment strategies for use in bioethanol production. Insect Science 17 (2010) 166-174.
- [47]M.E. Scharf, Z.J. Karl, A. Sethi, R. Sen, R. Raychoudhury, D.G. Boucias, Defining host-symbiont collaboration in termite lignocellulose digestion: "The view from the tip of the iceberg". Communicative & Integrative Biology 4 (2011) 761-763.
- [48]Y. Hongoh, Diversity and genomes of uncultured microbial symbionts in the termite gut. Bioscience, Biotechnology and Biochemistry 74 (2010) 1145-1151.
- [49]N. Shinzato, M. Muramatsu, T. Matsui, Y. Watanabe, Phylogenetic analysis of the gut bacterial microflora of the fungus-growing termite *Odontotermes formosanus*. Bioscience, Biotechnology and Biochemistry 71 (2007) 906-915.
- [50]J.A. Russell, C.S. Moreau, B. Goldman-Huertas, M. Fujiwara, D.J. Lohman, N.E. Pierce, Bacterial gut symbionts are tightly linked with the evolution of herbivory in

- ants. Proceedings of the National Academy of Sciences of the United States of America 106 (2009) 21236-21241.
- [51]W. Shi, S. Xie, X. Chen, S. Sun, X. Zhou, L. Liu, P. Gao, N.C. Kyrpides, E.G. No, J.S. Yuan, Comparative genomic analysis of the endosymbionts of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. PLoS genetics 9 (2013) e1003131.
- [52]D.W. Davidson, N.F. Anderson, S.C. Cook, C.R. Bernau, T.H. Jones, A.S. Kamariah, L.B. Lim, C.M. Chan, D.A. Clark, An Experimental Study of Microbial Nest Associates of Borneo's Exploding Ants (*Camponotus* [Colobopsis] species). Journal of Hymenoptera Research 18 (2009) 341-360.
- [53]J. Charles, Anatomie du gaster de la *Myrmica Rubra*. Étude sur les fourmis, les guêpes et les abeilles (1902) 8-26.
- [54]L. Hansen, W. Spangenberg, M. Galver, The infrabuccal chamber of *camponotus modoc* (Hymenoptera: Formicidae) - ingestion, digestion and survey of bacteria. Proceedings of the 3rd International Conference on Urban Pests, Spokane Falls Community College (1999) 211-219.
- [55]W. Wheeler, Ants - their structure, development and behavior. The Columbia University Press (1910) 13-39.
- [56]T. Eisner, W. Brown, The evolution and social significance of the ant proventriculus. (1958).
- [57]J.A. Gilbert, C.L. Dupont, Microbial metagenomics: beyond the genome. Annual Review of Marine Science 3 (2011) 347-371.
- [58]J. Handelsman, M.R. Rondon, S.F. Brady, J. Clardy, R.M. Goodman, Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & Biology 5 (1998) R245-249.
- [59]B. Temperton, S.J. Giovannoni, Metagenomics: microbial diversity through a scratched lens. Current Opinion in Microbiology 15 (2012) 605-612.
- [60]P.D. Schloss, J. Handelsman, Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biology 6 (2005) 229.
- [61]S. Abubucker, N. Segata, J. Goll, A.M. Schubert, J. Izard, B.L. Cantarel, B. Rodriguez-Mueller, J. Zucker, M. Thiagarajan, B. Henrissat, O. White, S.T. Kelley, B. Methe, P.D. Schloss, D. Gevers, M. Mitreva, C. Huttenhower, Metabolic

- reconstruction for metagenomic data and its application to the human microbiome. *PLoS Computational Biology* 8 (2012) e1002358.
- [62]F.D. Andreote, J.L. Azevedo, W.L. Araujo, Assessing the Diversity of Bacterial Communities Associated with Plants. *Brazilian Journal of Microbiology* 40 (2009) 417-432.
- [63]C. Manichanh, C.E. Chapple, L. Frangeul, K. Gloux, R. Guigo, J. Dore, A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. *Nucleic Acids Research* 36 (2008) 5180-5188.
- [64]L. Ranjard, F. Poly, J.C. Lata, C. Mougel, J. Thioulouse, S. Nazaret, Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: Biological and methodological variability. *Applied and Environmental Microbiology* 67 (2001) 4479-4487.
- [65]L. Ranjard, F. Poly, S. Nazaret, Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. *Research in Microbiology* 151 (2000) 167-177.
- [66]J. Goris, K.T. Konstantinidis, J.A. Klappenbach, T. Coenye, P. Vandamme, J.M. Tiedje, DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* 57 (2007) 81-91.
- [67]V. Torsvik, J. Goksoyr, F.L. Daae, High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology* 56 (1990) 782-787.
- [68]K. Nusslein, J.M. Tiedje, Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. *Applied and Environmental Microbiology* 65 (1999) 3622-3626.
- [69]N. Shah, H. Tang, T.G. Doak, Y. Ye, COMPARING BACTERIAL COMMUNITIES INFERRED FROM 16S rRNA GENE SEQUENCING AND SHOTGUN METAGENOMICS. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (2011) 165-176.
- [70]A. Buchan, S.Y. Newell, J.I. Moreta, M.A. Moran, Analysis of internal transcribed spacer (ITS) regions of rRNA genes in fungal communities in a southeastern U.S. salt marsh. *Microbial Ecology* 43 (2002) 329-340.

- [71]C.B. Blackwood, T. Marsh, S.H. Kim, E.A. Paul, Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. *Applied and Environmental Microbiology* 69 (2003) 926-932.
- [72]M.M. Fisher, E.W. Triplett, Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied and Environmental Microbiology* 65 (1999) 4630-4636.
- [73]R.B. Franklin, D.R. Taylor, A.L. Mills, Characterization of microbial communities using randomly amplified polymorphic DNA (RAPD). *Journal of Microbiological Methods* 35 (1999) 225-235.
- [74]F.B. Gich, E. Amer, J.B. Figueras, C.A. Abella, M.D. Balaguer, M. Poch, Assessment of microbial community structure changes by amplified ribosomal DNA restriction analysis (ARDRA). *International Microbiology: The Official Journal of the Spanish Society for Microbiology* 3 (2000) 103-106.
- [75]G. Muyzer, K. Smalla, Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek* 73 (1998) 127-141.
- [76]T. Thomas, J. Gilbert, F. Meyer, Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation* 2 (2012) 3.
- [77]J.F. Petrosino, S. Highlander, R.A. Luna, R.A. Gibbs, J. Versalovic, Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry* 55 (2009) 856-866.
- [78]H. Tamaki, C.L. Wright, X. Li, Q. Lin, C. Hwang, S. Wang, J. Thimmapuram, Y. Kamagata, W.T. Liu, Analysis of 16S rRNA amplicon sequencing options on the Roche/454 next-generation titanium sequencing platform. *Plos One* 6 (2011) e25263.
- [79]M. Ronaghi, Pyrosequencing sheds light on DNA sequencing. *Genome Research* 11 (2001) 3-11.
- [80]F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74 (1977) 5463-5467.
- [81]T. Tucker, M. Marra, J.M. Friedman, Massively parallel sequencing: the next big thing in genetic medicine. *American Journal of Human Genetics* 85 (2009) 142-154.

- [82]S. Kautz, B.E. Rubin, J.A. Russell, C.S. Moreau, Surveying the microbiome of ants: comparing 454 pyrosequencing with traditional methods to uncover bacterial diversity. *Applied and Environmental Microbiology* 79 (2013) 525-534.
- [83]Roche, Amplicon Fusion Primer Design Guidelines for GS FLX Titanium Series Lib-A Chemistry 454 Sequencing Technical Bulletin (2009).
- [84]Roche, GS FLX Titanium emPCR Method Manual (2009).
- [85]Roche, 454 Sequencing System Guidelines for Amplicon Experimental Design. (2011).
- [86]W.J. Ansorge, Next-generation DNA sequencing techniques. *New Biotechnology* 25 (2009) 195-203.
- [87]R.C. Novais, Y.R. Thorstenson, The evolution of Pyrosequencing for microbiology: From genes to genomes. *Journal of Microbiological Methods* 86 (2011) 1-7.
- [88]M. Kotik, Novel genes retrieved from environmental DNA by polymerase chain reaction: current genome-walking techniques for future metagenome applications. *Journal of Biotechnology* 144 (2009) 75-82.
- [89]H. Suenaga, Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environmental Microbiology* 14 (2012) 13-22.
- [90]X. Xiong, X. Yin, X. Pei, P. Jin, A. Zhang, Y. Li, W. Gong, Q. Wang, Retrieval of glycoside hydrolase family 9 cellulase genes from environmental DNA by metagenomic gene specific multi-primer PCR. *Biotechnology Letters* 34 (2012) 875-882.
- [91]S. Iwai, B. Chai, W.J. Sul, J.R. Cole, S.A. Hashsham, J.M. Tiedje, Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *The ISME Journal* 4 (2010) 279-285.
- [92]P.J.L. Bell, A. Sunna, M.D. Gibbs, N.C. Curach, H. Nevalainen, P.L. Bergquist, Prospecting for novel lipase genes using PCR. *Society for General Microbiology* 148 (2002) 2283-2291.
- [93]B.S. Kim, S.Y. Kim, J. Park, W. Park, K.Y. Hwang, Y.J. Yoon, W.K. Oh, B.Y. Kim, J.S. Ahn, Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library. *Journal of Applied Microbiology* 102 (2007) 1392-1400.
- [94]A. Labes, E.N. Karlsson, O.H. Fridjonsson, P. Turner, G.O. Hreggvidson, J.K. Kristjansson, O. Holst, P. Schonheit, Novel members of glycoside hydrolase family

- 13 derived from environmental DNA. *Applied and Environmental Microbiology* 74 (2008) 1914-1921.
- [95]K. Wong, T.T. Fong, K. Bibby, M. Molina, Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environment International* 45 (2012) 151-164.
- [96]J.R. Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, R.J. Farris, A.S. Kulam-Syed-Mohideen, D.M. McGarrell, T. Marsh, G.M. Garrity, J.M. Tiedje, The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* 37 (2009) D141-145.
- [97]S. Kumar, T. Carlsen, B.H. Mevik, P. Enger, R. Blaalid, K. Shalchian-Tabrizi, H. Kauserud, CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC Bioinformatics* 12 (2011) 182.
- [98]F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E.M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, R.A. Edwards, The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9 (2008) 386.
- [99]N. Mulder, R. Apweiler, InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods in Molecular Biology* 396 (2007) 59-70.
- [100]J.D. Thompson, T.J. Gibson, D.G. Higgins, Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* / editorial board, Andreas D. Baxevanis ... [et al.] Chapter 2 (2002) Unit 2 3.
- [101]R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32 (2004) 1792-1797.
- [102]R. Boyce, P. Chilana, T.M. Rose, iCODEHOP: a new interactive program for designing COnsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. *Nucleic Acids Research* 37 (2009) W222-228.
- [103]B. Contreras-Moreira, B. Sachman-Ruiz, I. Figueroa-Palacios, P. Vinuesa, primers4clades: a web server that uses phylogenetic trees to design lineage-specific PCR primers for metagenomic and diversity studies. *Nucleic Acids Research* 37 (2009) W95-W100.
- [104]QIAGEN, QIAamp® DNA Stool Mini Kit Handbook. For DNA purification from stool samples (2001).

- [105]AGENCOURT, AMPure® XP PCR Purification.
- [106]Invitrogen, Quant-iT™ PicoGreen® dsDNA Reagent and Kits (2008).
- [107]K. Biosystems, KAPA Library Quantification Kits. For Roche 454 GS FLX platform.
- [108]Roche, emPCR Method Manual – Lib-A LV. GS FLX Titanium Series (2009).
- [109]A. Morgulis, E.M. Gertz, A.A. Schaffer, R. Agarwala, A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 13 (2006) 1028-1040.
- [110]R.C. Edgar, B.J. Haas, J.C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27 (2011) 2194-2200.
- [111]R.C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26 (2010) 2460-2461.
- [112]P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G. Thallinger, D.J. Van Horn, C.F. Weber, Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75 (2009) 7537-7541.
- [113]J.D. Retief, Phylogenetic analysis using PHYLIP. *Methods in Molecular Biology* 132 (2000) 243-258.
- [114]M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn, The Pfam protein families database. *Nucleic Acids Research* 40 (2012) D290-301.
- [115]E.C. Dimmer, R.P. Huntley, Y. Alam-Faruque, T. Sawford, C. O'Donovan, M.J. Martin, B. Bely, P. Browne, W. Mun Chan, R. Eberhardt, M. Gardner, K. Laiho, D. Legge, M. Magrane, K. Pichler, D. Poggioli, H. Sehra, A. Auchincloss, K. Axelsen, M.C. Blatter, E. Boutet, S. Braconi-Quintaje, L. Breuza, A. Bridge, E. Coudert, A. Estreicher, L. Famiglietti, S. Ferro-Rojas, M. Feuermann, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, J. James, S. Jimenez, F. Jungo, G. Keller, P. Lemercier, D. Lieberherr, P. Masson, M. Moinat, I. Pedruzzi, S. Poux, C. Rivoire, B. Roechert, M. Schneider, A. Stutz, S. Sundaram, M. Tognolli, L. Bougueleret, G.

- Argoud-Puy, I. Cusin, P. Duek-Roggli, I. Xenarios, R. Apweiler, The UniProt-GO Annotation database in 2011. *Nucleic Acids Research* 40 (2012) D565-570.
- [116]R. Leinonen, R. Akhtar, E. Birney, L. Bower, A. Cerdeno-Tarraga, Y. Cheng, I. Cleland, N. Faruque, N. Goodgame, R. Gibson, G. Hoad, M. Jang, N. Pakseresht, S. Plaister, R. Radhakrishnan, K. Reddy, S. Sobhany, P. Ten Hoopen, R. Vaughan, V. Zalunin, G. Cochrane, The European Nucleotide Archive. *Nucleic Acids Research* 39 (2011) D28-31.
- [117]K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28 (2011) 2731-2739.
- [118]M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis, T.L. Madden, NCBI BLAST: a better web interface. *Nucleic Acids Research* 36 (2008) W5-9.
- [119]Y. Wurm, P. Uva, F. Ricci, J. Wang, S. Jemielity, C. Iseli, L. Falquet, L. Keller, Fourmidable: a database for ant genomics. *BMC Genomics* 10 (2009) 5.
- [120]Roche, Rapid Library Preparation Method Manual. GS FLX+ Series — XL+ (2011).
- [121]Roche, emPCR Method Manual – Lib-L LV. GS FLX Titanium Series (2009).
- [122]K.E. Anderson, J.A. Russell, C.S. Moreau, S. Kautz, K.E. Sullam, Y. Hu, U. Basinger, B.M. Mott, N. Buck, D.E. Wheeler, Highly similar microbial communities are shared among related and trophically similar ant species. *Molecular Ecology* 21 (2012) 2282-2296.
- [123]N.A. Moran, P. Tran, N.M. Gerardo, Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Applied and Environmental Microbiology* 71 (2005) 8802-8810.
- [124]F. Thomas, J.H. Hehemann, E. Rebuffet, M. Czjzek, G. Michel, Environmental and gut bacteroidetes: the food connection. *Frontiers in Microbiology* 2 (2011) 93.
- [125]X. Li, X. Nan, C. Wei, H. He, The gut bacteria associated with *Camponotus japonicus* Mayr with culture-dependent and DGGE methods. *Current Microbiology* 65 (2012) 610-616.
- [126]T. Kudo, Termite-microbe symbiotic system and its efficient degradation of lignocellulose. *Bioscience, Biotechnology and Biochemistry* 73 (2009) 2561-2567.
- [127]T. Kohler, C. Dietrich, R.H. Scheffrahn, A. Brune, High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the

- gut in wood-feeding higher termites (*Nasutitermes* spp.). Applied and Environmental Microbiology 78 (2012) 4691-4701.
- [128]E. Crotti, A. Rizzi, B. Chouaia, I. Ricci, G. Favia, A. Alma, L. Sacchi, K. Bourtzis, M. Mandrioli, A. Cherif, C. Bandi, D. Daffonchio, Acetic acid bacteria, newly emerging symbionts of insects. Applied and Environmental Microbiology 76 (2010) 6963-6970.
- [129]R. Reiss, J. Ihssen, M. Richter, E. Eichhorn, B. Schilling, L. Thony-Meyer, Laccase versus Laccase-Like Multi-Copper Oxidase: A Comparative Study of Similar Enzymes with Diverse Substrate Spectra. Plos One 8 (2013) e65633.
- [130]K.Y. Djoko, L.X. Chong, A.G. Wedd, Z. Xiao, Reaction mechanisms of the multicopper oxidase CueO from *Escherichia coli* support its functional role as a cuprous oxidase. Journal of the American Chemical Society 132 (2010) 2005-2015.
- [131]L. Martinkova, B. Uhnakova, M. Patek, J. Nesvera, V. Kren, Biodegradation potential of the genus *Rhodococcus*. Environment International 35 (2009) 162-177.
- [132]D.K. Ngugi, M.K. Tsanuo, H.I. Boga, *Rhodococcus opacus* strain RW, a resorcinol-degrading bacterium from the gut of *Macrotermes michaelseni*. African Journal of Biotechnology 4 (2005) 639-645.
- [133]S.D. Park, Y. Uh, I.H. Jang, K.J. Yoon, H.M. Kim, Y.J. Bae, *Rhodococcus erythropolis* septicaemia in a patient with acute lymphocytic leukaemia. Journal of Medical Microbiology 60 (2011) 252-255.
- [134]P.A. Grimont, F. Grimont, The genus *Serratia*. Annual Review of Microbiology 32 (1978) 221-248.